INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Unstable cycles in fluid Bramson networks

Vincent DUMAS

N° 2318
Août 1994

PROGRAMME 1

Rapport de recherche

Les rapports de recherche de l'INRIA
sont disponibles en format postscript sous
ftp.inria.fr (192.93.2.54)

si vous n'avez pas d'accès ftp
la forme papier peut être commandée par mail :
e-mail : dif.gesdif@inria.fr
(n'oubliez pas de mentionner votre adresse postale).

par courrier :
Centre de Diffusion
INRIA
BP 105 - 78153 Le Chesnay Cedex (FRANCE)

INRIA research reports
are available in postscript format
ftp.inria.fr (192.93.2.54)

if you haven't access by ftp
we recommend ordering them by e-mail :
e-mail : dif.gesdif@inria.fr
(don't forget to mention your postal address).

by mail :
Centre de Diffusion
INRIA
BP 105 - 78153 Le Chesnay Cedex (FRANCE)

CYCLES INSTABLES DANS LES RÉSEAUX DE BRAMSON FLUIDES.

Vincent Dumas[1]

Août 1994

Résumé.

Nous nous intéressons à une famille de réseaux de files d'attente à plusieurs classes de clients servis suivant la discipline FIFO. Ces réseaux comprennent deux files d'attente. Les clients pénètrent dans le réseau à la première file, puis vont à la deuxième file; ils y subissent J feed-backs successifs, avant de retourner à la première file, puis de quitter le réseau. Ce modèle a été introduit par Bramson [2] dans un cadre markovien; Bramson a montré que pour des valeurs de J très grandes et des intensités de trafic à chaque file très proches de 1 (mais inférieures), ces réseaux étaient instables (transients). C'était le premier exemple de réseau FIFO instable bien que les intensités de trafic soient plus petites que 1. Ce papier est consacré à l'analyse détaillée des modèles fluides associés aux réseaux de Bramson, et plus particulièrement à la preuve de l'instabilité fluide de ces réseaux pour $J \geq 2$ et pour un large spectre d'intensités de trafic plus petites que 1. Plus précisément, nous montrerons que le modèle fluide présente des "cycles instables": quand on part d'un état initial d'un type particulier, on revient ultérieurement à un état du même type mais de taille supérieure, et ainsi de suite. Notre analyse repose sur la notion de "couches homogènes", qui décrivent le mélange des différentes classes de clients à l'intérieur des files.

[1]INRIA, domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France. Vincent.Dumas@inria.fr

UNSTABLE CYCLES IN FLUID BRAMSON NETWORKS.

Vincent Dumas[2]

August 1994

Abstract.

This paper deals with a family of fluid queueing networks with several classes of customers and FIFO discipline. These networks consist of two queues. Customers enter the network at the first queue, and then go to the second queue; here they experience J successive feed-backs, and then they come back to the first queue, and finally they leave the network. This model was first presented by Bramson [2] in a Markovian setting; Bramson proved that these networks were unstable (transient) for very large values of J when the traffic intensities at each queue were chosen to be very close to 1 (but lower than 1). It was the first example of a queueing network that is unstable under FIFO discipline though the traffic intensities are smaller than 1. Our paper is devoted to the precise analysis of the fluid models associated to Bramson networks, and especially to the proof that fluid instability occurs for Bramson networks for $J \geq 2$ and for a wide range of traffic intensities lower than 1. More precisely, we will prove that the fluid model exhibits "unstable cycles": when it starts from a special initial state, it later goes back to a similar state with larger size, and then repeats this pattern infinitely often. Our analysis relies on the notion of "homogeneous layers", which describe the mixing of the different classes of customers inside the queues.

[2]INRIA, domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France. Vincent.Dumas@inria.fr

1 Introduction.

This paper deals with a family of fluid queueing networks with several classes of customers and FIFO discipline. The model of the network was first presented by Bramson [2]; it was the first example of a stochastic queueing network that is unstable under FIFO discipline though the traffic intensities at each queue are smaller than 1 (these are the usual stability conditions, which we will simply call "the usual conditions"). In order to prove this, he implicitly worked with the fluid model associated to his network, and proved that the unstable behaviour of this fluid model was followed by the original, stochastic model with positive probability.

The idea of analyzing complex, stochastic networks through fluid approximations is not new, but it recently gained new interest in the context of multiclass queueing networks, where the conditions of stability (i.e. the conditions of ergodicity of the state process) are generally not known. Initially Rybko and Stolyar [9] proved a kind of Foster criterion for Markov processes in discrete state spaces; this criterion of ergodicity involves a space-time scaling of the state process, thus leading to fluid approximations via functional laws of large numbers. Then Dai [3] extended this result to general state space, and he defined the fluid limit model associated to a stochastic queueing network. The initial Foster criterion implies that a stochastic queueing network is stable if the associated fluid network starting from some compact set empties in a finite time.

Notice that there is no converse of this criterion of stability, and neither is there an analogous criteria of instability (transience) in terms of the fluid model for the moment. However, in the few cases where a multiclass queueing network was proven to be unstable under the usual conditions, the associated fluid model exhibited "unstable cycles": when it started from a special initial state, it later went back to a similar state with larger size, and then repeated this pattern infinitely often. It must be emphasized that the usual conditions make each queue empty infinitely often, and then unstable cycles result in oscillations on the size of queues. This kind of behaviour was first discovered in deterministic models (see [8], [7]). Simple examples may be exhibited in the case where the state process is a homogeneous random walk in $\mathbb{Z}_+^N$ (for example when there is a discipline based on preemptive resume priorities between the classes, see [9], [1], [4]). Under some special conditions, this may be sufficient to prove the transience of the model (see [1]) by use of Malyshev and Menshikov's results on these random walks (see [11]). However, when the discipline is FIFO at every queue, the state space is much more complex and the fluid analysis is more difficult. To our knowledge, the only examples of unstable FIFO networks under the usual conditions were given by Seidman in [10] (in a deterministic setting) and by Bramson in [2]. Seidman exhibited unstable cycles in his model, but the specifity of FIFO is somewhat cancelled in his work since his unstable path works as if there were priorities between the classes. Our paper is devoted to the precise analysis of the fluid models associated to the class of networks introduced by Bramson in [2], and especially to the description of unstable cycles in these fluid models. We first briefly recall the characteristics of these networks and Bramson's results.

In these networks, customers arrive from outside at rate 1; they go to queue 1, where they require a service of mean δ, and then to queue 2 where they first require a service of mean c; then they experience J successive feed-backs at this queue, requiring at each stage a service of mean δ (see Figure 1). After that they come back to the first queue where they require a service of mean c, and then they leave the network. The external arrival process is Poisson and all the services are exponentially distributed and independent. The discipline is FIFO in both queues. Bramson's main result was that for some special values of the parameters c, J and δ, satisfying the usual conditions (which are: $c+\delta \leq c+J\delta < 1$), the network was unstable.

In his proof, Bramson implicitly referred to a fluid model of his network where δ and c^J are "infinitesimal", and: $1/2 < c < 1$. When this fluid model initially contains q customers at the second stage of the route (that is the first stage in queue 2), and no customer anywhere else, it later comes

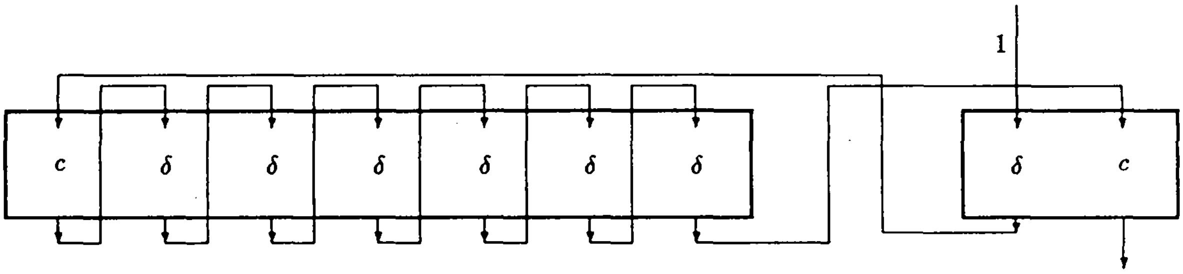

Figure 1 : Bramson network with $J = 6$.

back to approximately the same kind of state but with $\frac{c}{1-c}q$ $(> q)$ customers in queue 2. The analysis relies on the fact that periods when queue 1 remains empty alternate with periods when queue 2 remains empty, the transitions between two periods consisting of almost instantaneous transfers from one queue to the other one. The instantaneous transfers from queue 1 to queue 2 (resp. from queue 2 to queue 1) are caused by $\delta \simeq 0$ (resp. by $c^J \simeq 0$); they allow Bramson not to take care of the way the customers at different stages (i.e. the different "classes" of customers) are mixed inside the queues.

The essential part of Bramson's work then consists in controlling the deviation of the stochastic model from this unstable, fluid behaviour. His approach finally leads to imposing that c be very close to 1 (larger than $1 - 1/400$), δ be very close to 0, and above all J be very large (of the order of 1600). Notice however that the fluid part of his analysis already imposed to take J very large (because c^J was assumed to be infinitesimal) so that he did not have to study the mixing of the classes.

In this paper, we will prove that fluid instability occurs for Bramson networks under the usual conditions even for low values of J. More precisely, for general values of J, we will consider the associated fluid model with: $c \in]0,1[$, and: $\delta = 0$; then the usual conditions will be satisfied, and it is easy to check that the network is stable if $c < 1/2$. When the initial state has a special form and c belongs to some special interval (included in $]1/2, 1[$), we will be able to describe the exact path followed by the fluid model; after some time, the network exhibits the same, special form as initially, and for $J \geq 3$, there are values of c for which this cycle is unstable (the new state is bigger than the former one); moreover, the set of these values is an interval which is growing with J and is converging to $]-1/2, 1[$ when J tends to $+\infty$. Because of the mixing of the different classes inside the queues, there will never be instantaneous transfers of customers (those with null services will always be slowed by those with non-null services). This mixing will be described in terms of "homogeneous layers"; this essential notion will be the basis for our whole analysis. It is equivalent to the piecewise linearity of the departure processes of the different classes.

We will then have described exact, unstable cycles of the fluid model for low values of J (in fact they exist for $J \geq 2$); these unstable cycles are globally similar to those already encountered in networks with priorities, but they are specific to the FIFO discipline, which appears through the existence of homogeneous layers.

In the second section, we will recall the definition of the fluid limit model by Dai, give its characteristics in the FIFO case, introduce the notion of homogeneous layers and prove some general, basic results about FIFO multiclass fluid networks. The third section is entirely devoted to fluid Bramson networks and especially to the proof of the main result of the paper (Theorem 3.4). In conclusion we will make some comments about open problems concerning fluid limit models and especially future work to achieve on Bramson networks.

Notations: we will use the notation $a \wedge b$ for $\min(a, b)$, $a \vee b$ for $\max(a, b)$, and $[a]^+$ for $a \vee 0$.

2 Fluid models with FIFO discipline: characteristics and basic results.

2.1 Fluid limit models.

Most of this paragraph comes from [6]. Let us consider a network consisting of K queues (denoted by an index $k = 1, ..., K$) and I types of customers (indexed by $i = 1, ..., I$). At each queue there is a waiting room of infinite capacity and a non-idling server working at unit-speed. Type i customers follow a fixed route of length l_i through the network, i.e. they visit l_i queues and the queue visited at the s^{th} stage of their route ($1 \leq s \leq l_i$) is k_{is} (and so they enter the network at queue k_{i1} and leave it at queue k_{il_i}). A customer of type i at stage s of his route is said to belong to the class (i, s).

The external arrivals of type i customers are supposed to form a Poisson process of intensity ν_i. At each visited queue, these customers require some service time; the service times required by class (i, s) customers are supposed to be independent and exponentially distributed with mean $\frac{1}{\mu_{is}} > 0$. The different sequences or processes are assumed to be independent. At last, the service discipline is FIFO at each queue (customers are served in their order of arrivals, regardless of their classes).

A markovian description of the state of the network at time t is given by the variable:

$$X_t = (C_t(k, l))_{\substack{1 \leq k \leq K \\ 1 \leq l \leq Q_t(k)}}$$

where:

- $Q_t(k)$ is the number of customers at queue k at time t;
- $C_t(k, l)$ is the class of the l^{th} customer in queue k at time t (customers are supposed to be ordered in the queue according to their order of arrival).

It is easy to verify that $(X_t)_{t\geq 0}$ is a homogeneous Markov process with countable state space. As usual, we identify the stability of our network with the ergodicity of $(X_t)_{t\geq 0}$. The conditions of stability will depend on the traffic intensities of the different classes.

Definition 2.1
Let us denote : $\rho_{is} = \nu_i/\mu_{is}$, $\forall (i, s)$ *;* ρ_{is} *is the traffic intensity of class* (i, s) *customers at queue* $k = k_{is}$*. The traffic intensity at queue* k *is defined by :*

$$\rho_k = \sum_{k_{is}=k} \rho_{is}.$$

If the network is stable, ρ_{is} is the average amount of work brought by the class (i, s) for the server k_{is}. It is then obvious (and easy to prove) that a general necessary condition for the stability of our network is that the traffic intensities at each queue are less than 1, that is:

$$\rho_k < 1, \quad \forall k.$$

We will call these conditions "the usual conditions", since they are the exact stability conditions for "classic" networks. Bramson gave an example proving that it is not true in general (some special networks may be unstable under the usual conditions). We will analyze the fluid version of his model in this paper.

Let us now present the notion of fluid limit model introduced by Dai ([3]). For this we need a few more definitions and notations.

Definition 2.2
For a given initial state x and a given class (i,s):

- $Q_t^x(i,s)$ *is the number of class (i,s) customers at time t ;*
- $A_t^x(i,s)$ *is the number of class (i,s) arrivals up to time t (with the convention : $A_0^x(i,s) = Q_0^x(i,s)$);*
- $D_t^x(i,s)$ *is the number of class (i,s) departures up to time t (with the convention: $D_0^x(i,s) = 0$);*
- $W_t^x(i,s)$ *is the load (or the work time) constituted by class (i,s) customers at time t ;*
- $\Omega_t^x(i,s)$ *is the total load brought by class (i,s) customers up to time t (with the convention: $\Omega_0^x(i,s) = W_0^x(i,s)$);*
- $B_t^x(i,s)$ *is the time spent by server k_{is} to serve class (i,s) customers up to time t (with the convention: $B_0^x(i,s) = 0$).*

All these processes are taken right-continuous.

To all these class processes in the form : $H^x(i,s) = (H_t^x(i,s))_{t\geq 0}$, we may associate the queue processes :

$$H^x(k) \triangleq \sum_{k_{is}=k} H^x(i,s), \quad 1 \leq k \leq K, \quad \text{and the vector processes:}$$

$$H^x \triangleq (H^x(i,s))_{\substack{1\leq i\leq I \\ 1\leq s\leq l_i}} .$$

Now it is time we defined the fluid limit model. Let $\|\ \|$ be a norm on the vector space $\mathbb{R}^C$ (where $C = \sum_{i=1}^{I} l_i$ is the total number of classes), and let x be a state of our network. We define :

$$f(x) = \|(q(i,s))_{\substack{1\leq i\leq I \\ 1\leq s\leq l_i}}\|,$$

where $q(i,s)$ is the number of class (i,s) customers in state x.

For any sequence (x_n) with $f(x_n) > 0, \forall n$, and any vector process H^{x_n}, we define the scaled version $\overline{H}^n$ of this process by :

$$\forall t \geq 0 : \quad \overline{H}_t^n = \frac{H^{x_n}_{f(x_n)t}}{f(x_n)}.$$

By definition, we have:

$$\forall n : \quad \|\overline{Q}_0^n\| = \frac{\|(Q_0^{x_n}(i,s))_{\substack{1\leq i\leq I \\ 1\leq s\leq l_i}}\|}{f(x_n)} = 1.$$

The following result was shown in [3]. It is a simple consequence of the relative compactness of the sequence $(\overline{B}^n)$ and of functional laws of large numbers.

Definition 2.3
For all $q \in \mathbb{R}^C$ with $\|q\| = 1$, and any sequence (x_n) with $f(x_n) \to +\infty$ and $\overline{Q}_0^n \to q$, there is a subsequence of $(\overline{Q}^n, \overline{A}^n, \overline{D}^n, \overline{W}^n, \overline{\Omega}^n, \overline{B}^n)$ that converges in distribution to a limit (Q, A, D, W, Ω, B). Any particular limit will be called a fluid limit *of the network with "initial state" q. The set of all the fluid limits forms the* fluid limit model.

Remark 2.4
In view of the markovian state process X_t, it is obvious that q only partially describes the fluid, initial state of the network, which accounts for the quotation used in the above definition. We will later give a complete description of a fluid state of the network.

For any fluid limit, the limit process B is continuous, and then all the limit processes are continuous, since they derive from B by the following relations: for any class (i, s) and all $t \geq 0$:

$$Q_t(i,s) = A_t(i,s) - D_t(i,s), \text{ with : } A_0(i,s) = Q_0(i,s) = q_{is}, D_0(i,s) = 0; \quad (1)$$

$$A_t(i,s) = Q_0(i,s) + D_t(i,s-1), \text{ with : } D_t(i,0) \triangleq \nu_i t; \quad (2)$$

$$\Omega_t(i,s) = \frac{A_t(i,s)}{\mu_{is}}; \quad (3)$$

$$B_t(i,s) = \frac{D_t(i,s)}{\mu_{is}}; \quad (4)$$

$$W_t(i,s) = \Omega_t(i,s) - B_t(i,s). \quad (5)$$

Two additional relations characterize the service discipline (see [6] for a justification):

Conservativeness: this relation describes how the arriving load is processed in each queue regardless of the ordering of customers.

$$\forall k, \forall t \geq 0: \quad B_t(k) = \inf_{0 \leq s \leq t} (\Omega_s(k) + t - s) \wedge t; \quad (6)$$

FIFO: this relation conveys the fact that the waiting time of a customer is the load he finds in the queue at his arrival.

$$\forall (i,s), k_{is} = k, \forall t \geq 0: \quad D_{t+W_t(k)}(i,s) = A_t(i,s). \quad (7)$$

We will be interested by deterministic functions (Q, A, D, W, Ω, B) satisfying all the above relations. If (Q, A, D, W, Ω, B) belongs to the support of a fluid limit, then it satisfies these relations, but notice that the converse may not be true. However, when we will have given a complete description of a fluid state in the fluid limit model, if for a given initial state there is only one set of (deterministic) functions satisfying all these equations, then it will be a fluid limit of the network.

The notion of fluid limit model may be extended to a wider category of networks and to more general assumptions on the laws of the variables (the services of one class will simply be i.i.d. and its external arrivals will form a renewal process with unbounded and spread-out interarrival times; see [3]) for which there is still a markovian description of the state. The main result connecting the fluid limit model with the original, stochastic model is then Theorem 4.3 of [3], generalizing a result of [9].

Theorem 2.5
Assume that a given network is such that there exists a constant $T > 0$ which satisfies:

$$\forall t \geq T, \quad \forall (i,s): \quad Q_t(i,s) = 0 \quad \textit{(or equivalently:} \quad W_t(i,s) = 0\textit{)},$$

whatever the particular fluid limit considered. Then the original model is (Harris) ergodic.

We will make use of this theorem to prove an easy result at the beginning of section 3. We are now going to show how the fluid model works in some simple cases.

2.2 Basic results for a multiclass simple queue and a simple queue with multiple feed-backs.

We will now consider the fluid limit model of some network, and we will concentrate on a single queue inside this network. We assume that there are N customer classes at this queue, denoted by an index n, $1 \leq n \leq N$. The notations will thus slightly differ from those previously defined: exceptionally (and only in this section), the processes with an index will be class processes refering to the corresponding class, and those without index will be queue processes refering to the queue (no vector process will be used). As far as we concentrate on this queue, only the following relations (which directly derive from the previous ones) will be needed; for all $t \geq 0$:

$$\Omega_t = \sum_{n=1}^{N} \frac{A_t(n)}{\mu_n}, \tag{8}$$

$$B_t = \sum_{n=1}^{N} \frac{D_t(n)}{\mu_n}, \tag{9}$$

$$W_t = \Omega_t - B_t, \tag{10}$$

$$\forall n: \quad Q_t(n) = A_t(n) - D_t(n), \tag{2}$$

and:

$$B_t = \inf_{0 \leq s \leq t} (\Omega_s + t - s) \wedge t, \tag{7}$$

$$\forall n: \quad D_{t+W_t}(n) = A_t(n). \tag{8}$$

Consider processes $A(n)$, $1 \leq n \leq N$, as the input to the queue, and processes $D(n)$, $1 \leq n \leq N$, as the output from the queue. The first question is: how to derive the output from the input ? As: $t \mapsto t + W_t$, is a continuous, non-decreasing function (the load is processed at unit-speed), mapping $[0, +\infty[$ into $[W_0, +\infty[$, it is easy to check that the above sequence of equations allows us to derive the output on $[W_0, t + W_t[$ from the input on $[0, t[$. Since the discipline is FIFO, the departure processes on $[0, W_0[$ cannot be calculated; they are determined by the initial state of the queue, that is, by the densities of the respective classes inside the queue. In fact, we may consider that the fluid state of the queue at any time t is given by W_t and: $(D_u(n) - D_t(n))_{t<u\leq t+W_t}$, for $1 \leq n \leq N$. This important point will later be crucial to understand the definition of homogeneous layers.

We recall that A, D, Ω, B are continuous, non-decreasing functions. At any time t when a function H of the fluid model admits a derivative (that is at almost every time, see [4]), we will denote it by $\dot{H}_t$, and of course when we will use this expression, it will implicitly imply that the derivative exists.

Lemma 2.6
Assume that at some $t \geq 0$, we have $W_t > 0$. Then:

$$\dot{B}_t = 1.$$

More, if $\dot{A}_t(n)$ exists, $1 \leq n \leq N$, and $\dot{\Omega}_t = \sum_{n=1}^{N} \frac{\dot{A}_t(n)}{\mu_n} > 0$, then:

$$\dot{D}_{t+W_t}(n) = \frac{\dot{A}_t(n)}{\dot{\Omega}_t}, \quad 1 \leq n \leq N.$$

Proof :
For $u \in [t, t+W_t]$, we have:

$$\begin{aligned} B_u &= \inf_{0 \leq s \leq u} (\Omega_s + u - s) \wedge u \\ &= \inf_{0 \leq s \leq t} (\Omega_s + t - s + u - t) \wedge (t + u - t) \wedge (\inf_{t \leq s \leq u} (\Omega_s + u - s) \\ &= (B_t + u - t) \wedge (\inf_{t \leq s \leq u} (\Omega_s + u - s). \end{aligned}$$

But for $s \in [t, u]$:

$$\Omega_s + u - s \geq \Omega_t = B_t + W_t \geq B_t + u - t,$$

and consequently:

$$B_u - B_t = u - t.$$

Thus B_t admits a right derivative equal to 1 as long as $W_t > 0$, and then also a left derivative equal to 1 since W_t is continuous.

Now, as we already noticed, $s \mapsto s + W_s$ is a continuous, non-decreasing function; moreover, its derivative at time t is:

$$1 + \dot{W}_t = 1 + \dot{\Omega}_t - \dot{B}_t = \dot{\Omega}_t > 0.$$

So in view of relation (7), for any class n we are in the following situation: we have two continuous functions $f = D(n)$ and $g = I + W$ (I denotes the identity map), mapping $\mathbb{R}^+$ into itself; g is non-decreasing; at time t, g and $f \circ g = A(n)$ admit derivatives, with $\dot{g}(t) > 0$. Then it is easy to check that f admits a derivative at time $g(t)$, and that: $\dot{f}(g(t)) = \dfrac{\dot{\overline{f \circ g}}(t)}{\dot{g}(t)}$. The conclusion is then immediate. □

In fact, we will only consider situations where the evolution of the network is piecewise linear. So let us examine more accurately what kind of phenomena hide behind equation (6) in this case.

Lemma 2.7
Assume that at time $t_0 \geq 0$, we have: $W_{t_0} = w \geq 0$, and: $\dot{\Omega}_t = \rho \geq 0$ for $t \in]t_0, t_0 + \delta[$ ($\delta > 0$). Then, for $t \in [t_0, t_0 + \delta]$:

$$W_t = [w + (\rho - 1)(t - t_0)]^+.$$

Proof :
For $t \in [t_0, t_0 + \delta]$, we have:

$$B_t = [B_{t_0} + t - t_0] \wedge \inf_{t_0 \leq s \leq t} [\Omega_s + t - s],$$

or equivalently:

$$\begin{aligned} W_t &= \Omega_t - B_t \\ &= [W_{t_0} + (\Omega_t - \Omega_{t_0}) - (t - t_0)] \vee \sup_{t_0 \leq s \leq t} [(\Omega_t - \Omega_s) - (t - s)] \\ &= [w + (\rho - 1)(t - t_0)] \vee \sup_{t_0 \leq s \leq t} [(\rho - 1)(t - s)] \\ &= [w + (\rho - 1)(t - t_0)] \vee [(\rho - 1)(t - t_0)] \vee 0 \\ &= [w + (\rho - 1)(t - t_0)]^+. \end{aligned}$$

□

Notice that this result does not involve the FIFO discipline nor the different classes of customers (it is true for any work-conserving queue). Approximately speaking, it tells us that when $W_t = 0$, then $\dot{B}_t = \dot{\Omega}_t \wedge 1$ (or $\dot{W}_t = [\dot{\Omega}_t - 1]^+$) (whereras when $W_t > 0$, Lemma 2.6 taught us that: $\dot{W}_t = \dot{\Omega}_t - 1$).

From now on, we will only have to consider piecewise linear processes, because the states of the queues will always consist of the superposition of "homogeneous layers". We will soon give a precise definition to this notion, but let us explain its intuitive meaning first.

At the stochastic level, the description of the state of a given queue consists in listing the customers who are waiting for service from the first to the last (in the order of arrival) and specifying the class to which each of them belong. If there is a lot of customers in the queue and if you have a global view over this queue, you will not see individual customers but densities of classes all along the queue. If in some portion of this queue the respective densities are constant, we might say that this portion (or this *layer* if you consider that customers accumulate at the "top" of the queue) is *homogeneous* in the sense that classes are homogeneously mixed inside it.

These notions are in fact easier to define at the fluid level. Consider a fluid queue at some time t with a positive amount of customers, that is: $W_t > 0$. These customers now actually form a continuum, and we cannot any longer identify a portion of the queue by telling the ranks of the first customer and the last customer of this portion; but we have a new measure of the rank in the queue which is given by the instant of service $u \in]t, t + W_t]$: the beginning (or the *bottom*) of the queue corresponds to time t and the end (or the *top*) corresponds to time $t + W_t$. A *layer* will then be any subinterval of $]t, t + W_t]$. As we already noticed, the list of the classes to which the customers belong, from the first to the last, is replaced in the fluid context by the departure processes of the different classes, described from time t until time $t + W_t$ (that is: $(D_u(n) - D_t(n))_{t<u\leq t+W_t}$, for $1 \leq n \leq N$). Since constant densities of the respective classes inside the queue result in constant departure rates, the homogeneity of a layer will then be defined in terms of constant departure rates on the corresponding subinterval of $]t, t + W_t]$.

Notice however that these departure rates must satisfy a constraint which is implied by Lemma 2.6 and relation (9): at any time u where derivatives exist and $W_u > 0$, we must have:

$$\sum_{n=1}^{N} \frac{\dot{D}_u(n)}{\mu_n} = 1. \tag{12}$$

These considerations are summarized and formalized in the following definition.

Definition 2.8

- *Assume that $W_t > 0$. The fluids present in the queue at time t will be processed during the time interval $]t, t + W_t]$. Any part of these fluids that will be processed during a time interval $]t + w_1, t +$*

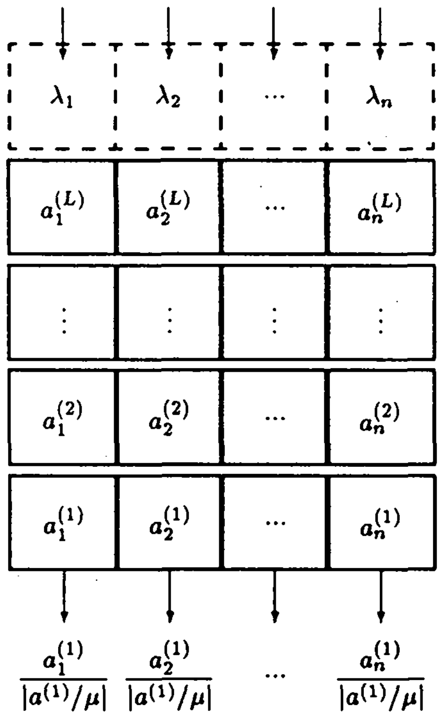

Figure 2 : Superposition of homogeneous layers.

$w_1 + w_2] \subseteq]t, t + W_t]$ is called a layer of thickness w_2. The state of the queue may be arbitrarily considered as the superposition *of the layers $]t + \sum_{l'<l} w^{(l')}, t + \sum_{l'\leq l} w^{(l')}]$, $1 \leq l \leq L$, for any sequence $(w^{(l)})_{1\leq l\leq L}$ such that $W_t = \sum_{l=1}^{L} w^{(l)}$.*

- *Let $a = [a_1, a_2, ..., a_N]$ be a non-negative, non-null vector of length N. Denote by a/μ the vector of components a_n/μ_n, $1 \leq n \leq N$, and $|v| = v_1 + ... + v_N$ for any vector $v = [v_1, ..., v_N]$. We say that the layer $]t + w_1, t + w_1 + w_2] \subseteq]t, t + W_t]$ is a homogeneous layer of composition a (and thickness w_2) if:*

$$\forall u \in]t + w_1, t + w_1 + w_2[: \quad \dot{D}_u(n) = \frac{a_n}{|a/\mu|}, \quad 1 \leq n \leq N. \tag{13}$$

Of course the composition vector a is defined up to a positive, multiplicative constant. It gives the proportions of the different classes in the layer $]t + w_1, t + w_1 + w_2]$. Indeed, the amount of class n customers in this layer (which we will denote $q(n)$) is the quantity of class n customers who will be served during this time interval, and then:

$$q(n) = D_{t+w_1+w_2} - D_{t+w_1}(n) = \frac{a_n}{|a/\mu|} w_2,$$

and consequently:

$$q(n) = \frac{a_n}{|a|} q,$$

where q is the total amount of customers in this layer. Notice also that the thickness of a homogeneous layer of composition a may be determined by the amount of any class.

A superposition of L homogeneous layers will be represented as in Figure 2. Customers of different classes arrive at to the queue at the top of the superposition, and leave it from the bottom. Each row (resp. each column) corresponds to a layer (resp. to a class), and in each case picturing a class in a layer is written its "proportion" (up to a positive scalar). Arrows symbolize the flows of different classes, with the corresponding rates written aside. The dashed lines represent the top of the queue, where a new layer will be created by the arrivals of rates λ_n, $1 \leq n \leq N$. Indeed, a homogeneous layer is produced by constant arrival rates, as stated in the following proposition.

Proposition 2.9
Assume that at time $t_0 \geq 0$, we have $W_{t_0} = w \geq 0$, and there is a non-negative, non-null vector $\lambda = [\lambda_1, ..., \lambda_N]$ such that for some $\delta > 0$ and all n, $1 \leq n \leq N$:

$$\dot{A}_u(n) = \lambda_n \quad \text{for } u \in]t_0, t_0 + \delta[.$$

Denote by $\rho = \sum_{n=1}^{N} \frac{\lambda_n}{\mu_n}$ the traffic intensity on the time interval $]t, t+\delta[$. Then, for $t \in [t_0, t_0 + \delta]$:

$$W_t = [w + (\rho - 1)(t - t_0)]^+,$$

and as long as $W_t > 0$, the part of fluid above the original state (analytically, it is the layer $](t_0 + W_{t_0}) \vee t, t + W_t])$ is a homogeneous layer of composition λ (and thickness $(W_t - [W_{t_0} - (t - t_0)]^+)$).

Proof :
The first part of the proposition is a direct application of lemma 2.7, since $\dot{\Omega}_t = \rho$ for $t \in]t_0, t_0 + \delta[$ in view of (8). The thickness of the above layer is:

$$t + W_t - (t_0 + W_{t_0}) \vee t = W_t - [W_{t_0} - (t - t_0)]^+.$$

Then, assume that at time $t \in]t_0, t_0 + \delta[$ we have: $W_t = [w + (\rho - 1)(t - t_0)]^+ > 0$, and thus for $s \in]t_0, t]$:

$$W_s = [w + (\rho - 1)(s - t_0)]^+ > 0.$$

In consequence, by lemma 2.6, for $s \in]t_0, t]$:

$$\forall n : \quad \dot{D}_{s+W_s}(n) = \frac{\lambda}{\rho},$$

and especially for $u \in](t_0 + W_0) \vee t, t + W_t[$:

$$\forall n : \quad \dot{D}_u(n) = \frac{\lambda}{|\lambda/\mu|},$$

which exactly means that the above layer has composition λ. □

Now assume that the queue consists of $N - 1$ successive feed-backs; analytically:

$$\forall t \geq 0 : \quad A_t(n) = Q_0(n) + D_t(n-1), \quad 1 < n \leq N.$$

In order to analyze this queue, we must come back to the stochastic network for a short time and define a special process.

Definition 2.10
Let $\mathbf{W}_t^x$ be the potential load of the queue at time t for an initial condition x, that is the cumulated service time required by the customers present at the queue at time t, including all their expected feed-backs. We have:

$$\mathbf{W}_t^x = \Omega_t^x - B_t^x,$$

with Ω_t^x being the cumulated load brought to the queue up to time t (each customer arriving at the queue brings the sum of his successive service times).

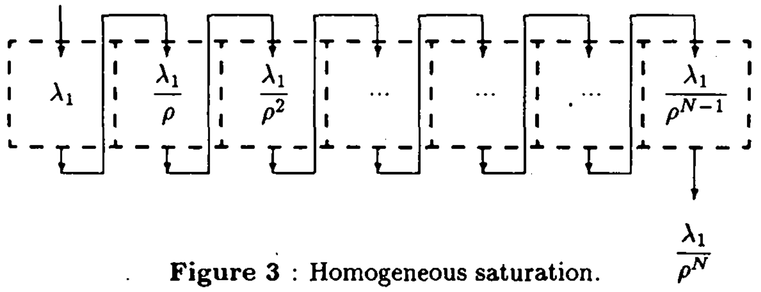

Figure 3 : Homogeneous saturation.

It is obvious that the busy periods of the queue would be the same if the server performed all the successive services of each customer in one time; in this case, $\boldsymbol{\Omega}_t^x$ would be the usual load brought to the queue up to time t, and consequently:

$$\forall t \geq 0: \quad B_t^x = \inf_{0 \leq s \leq t} [\boldsymbol{\Omega}_s^x + t - s] \wedge t.$$

More, it is easy to check that:

$$\begin{aligned} \overline{\boldsymbol{W}}^n \to \boldsymbol{W} &\triangleq \sum_{n=1}^{N} \frac{\sum_{p \leq n} Q(p)}{\mu_n} \\ &= \sum_{n=1}^{N} \frac{\boldsymbol{Q}(n)}{\mu_n}, \end{aligned}$$

with $\boldsymbol{Q}(n) \triangleq \sum_{p \leq n} Q(p)$.

Assume that at time $t_0 \geq 0$, we have $\boldsymbol{W}_{t_0} = w \geq 0$, and there is a constant $\lambda_1 > 0$ such that for some $\delta > 0$:

$$\dot{A}_t(1) = \lambda_1 \quad \text{for } t \in]t_0, t_0 + \delta[.$$

Denote by

$$\rho = \sum_{n=1}^{N} \frac{\lambda_1}{\mu_n}$$

what we will call the potential traffic intensity on the time interval $]t_0, t_0 + \delta[$. We will prove the following result.

Proposition 2.11

- *Assume that $\rho < 1$ and $w \leq (1-\rho)\delta$. Then $W_t > 0$ for $t \in [t_0, t_0 + \frac{w}{1-\rho}[$ and $W_t = 0$ for $t \in [t_0 + \frac{w}{1-\rho}, t_0 + \delta]$.*
- *Assume that $\rho > 1$ and $w = 0$ (the queue is initially empty). Let $f(x) \triangleq \sum_{n=1}^{N} \frac{\lambda_1}{x^{n-1}\mu_n}$, and let $\rho > 1$ be the unique positive solution of: $f(x) = x$, or $\sum_{n=1}^{N} \frac{\lambda_1}{x^n \mu_n} = 1$. Assume that $\sum_{n=1}^{N-1} \frac{n\lambda_1}{\mu_{n+1}} < 1$. Then for $t \in]t_0, t_0 + \delta[$ we have: $\dot{W}_t = \rho - 1$, and the state is a homogeneous layer of composition $[\frac{\lambda_1}{\rho}, \frac{\lambda_1}{\rho^2}, \ldots, \frac{\lambda_1}{\rho^N}]$. This evolution is described by figure 3 (it depicts the situation at time t_0^+).*

Proof :
Since $\dot{\boldsymbol{\Omega}}_t = \rho$ for $t \in]t_0, t_0 + \delta[$, and $B_t = \inf_{0 \leq s \leq t} [\boldsymbol{\Omega}_s + t - s] \wedge t$, lemma 2.7 gives us:

$$\boldsymbol{W}_t = [w + (\rho - 1)(t - t_0)]^+ \quad \text{for } t \in [t_0, t_0 + \delta].$$

- Since: $(W_t > 0) \Leftrightarrow (\boldsymbol{W}_t > 0)$, the result for $\rho < 1$ results immediately from this formula.
- If $\rho > 1$, then $W_t > 0$ for $t \in]t_0, t_0 + \delta[$. In consequence:

$$\dot{B}_t = 1 \quad \text{for } t \in]t_0, t_0+\delta[, \text{ and then:}$$

$$W_t - W_{t_0} = (\Omega_t - \Omega_{t_0}) - (B_t - B_{t_0}) \quad \text{results in:}$$

$$W_t = (\Omega_t - \Omega_{t_0}) - (t - t_0) \quad \text{for } t \in]t_0, t_0 + \delta[. \tag{14}$$

Consider the sequence:

$$\begin{cases} \rho_0 = \rho \\ \rho_{i+1} = f(\rho_i) \end{cases}$$

Since f is a convex, decreasing function, satisfying: $f(1) = \rho > 1$, and: $\dot{f}(1) = -\sum_{n=1}^{N-1} \frac{n\lambda_1}{\mu_{n+1}} > -1$, it is easy to check that the sequence (ρ_i) takes its values in $]1, +\infty[$, and that it tends to ρ when i tends to $+\infty$.

We will prove that the following sequence of implications holds: for $i \in \mathbb{N}$:

$$\begin{aligned}
&(W_t \leq (\rho_{2i} - 1)(t - t_0) \quad \text{for } t \in [t_0, t_0 + \delta]) \\
&\overset{1}{\Longrightarrow} \left(\forall n : D_t(n) - D_{t_0}(n) \geq \frac{\lambda_1}{\rho_{2i}^n}(t - t_0) \quad \text{for } t \in [t_0, t_0 + \rho_{2i}\delta]\right) \\
&\overset{2}{\Longrightarrow} (W_t \geq (\rho_{2i+1} - 1)(t - t_0) \quad \text{for } t \in [t_0, t_0 + \delta]) \\
&\overset{3}{\Longrightarrow} \left(\forall n : D_t(n) - D_{t_0}(n) \leq \frac{\lambda_1}{\rho_{2i+1}^n}(t - t_0) \quad \text{for } t \in [t_0, t_0 + \rho_{2i+1}\delta]\right) \\
&\overset{4}{\Longrightarrow} (W_t \leq (\rho_{2i+2} - 1)(t - t_0) \quad \text{for } t \in [t_0, t_0 + \delta])
\end{aligned}$$

At first we have for $t \in [t_0, t_0 + \delta]$:

$$W_t \leq \boldsymbol{W}_t = (\rho - 1)(t - t_0) = (\rho_0 - 1)(t - t_0),$$

so that the first inequality for $i = 0$ is satisfied. Now assume that for $t \in [t_0, t_0 + \delta]$ and some $i \in \mathbb{N}$:

$$W_t \leq (\rho_{2i} - 1)(t - t_0).$$

For $t \in [t_0, t_0 + \delta]$ and all n, we get:

$$D_{t+(\rho_{2i}-1)(t-t_0)}(n) \geq D_{t+W_t}(n) = A_t(n).$$

Since $Q_{t_0}(n) = A_{t_0}(n) - D_{t_0}(n) = 0$, it is equivalent to say that for $t \in [t_0, t_0 + \rho_{2i}\delta]$:

$$\begin{aligned}
D_t(n) - D_{t_0}(n) &\geq A_{t_0 + \frac{t-t_0}{\rho_{2i}}}(n) - A_{t_0}(n) \\
&= D_{t_0 + \frac{t-t_0}{\rho_{2i}}}(n-1) - D_{t_0}(n-1).
\end{aligned}$$

Since $\rho_{2i} > 1$, by an immediate induction on n we get that for $t \in [t_0, t_0 + \rho_{2i}\delta]$:

$$\begin{aligned}
D_t(n) - D_{t_0}(n) &\geq A_{t_0 + \frac{t-t_0}{\rho_{2i}^n}}(1) - A_{t_0}(1) \\
&= \frac{\lambda_1}{\rho_{2i}^n}(t - t_0).
\end{aligned}$$

Thus we proved $\overset{1}{\Longrightarrow}$.

Since the departure rate of class $(n-1)$ customers is the arrival rate of class n customers, we may equivalently write that for $t \in [t_0, t_0 + \rho_{2i}\delta]$ and all n (including $n = 1$):

$$A_t(n) - A_{t_0}(n) \geq \frac{\lambda_1}{\rho_{2i}^{n-1}}(t - t_0).$$

In view of (8), we deduce that for $t \in [t_0, t_0 + \rho_{2i}\delta]$, we have:

$$\Omega_t - \Omega_{t_0} \geq \sum_{n=1}^{N} \frac{\lambda_1}{\rho_{2i}^{n-1}\mu_n}(t - t_0) = f(\rho_{2i})(t - t_0) = \rho_{2i+1}(t - t_0).$$

By a simple application of (14), we deduce that for $t \in [t_0, t_0 + \delta]$:

$$W_t \geq (\rho_{2i+1} - 1)(t - t_0),$$

and thus we proved $\stackrel{2}{\Longrightarrow}$.

It is easy to check that $\stackrel{3}{\Longrightarrow}$ (resp. $\stackrel{4}{\Longrightarrow}$) may be proven by the same method as for $\stackrel{1}{\Longrightarrow}$ (resp. $\stackrel{2}{\Longrightarrow}$). In consequence, all these inequalities are valid for any $i \in \mathbb{N}$. By letting i tend to $+\infty$, we thus obtain that for $t \in [t_0, t_0 + \delta]$:

$$W_t = (\rho - 1)(t - t_0),$$

and for $t \in [t_0, t_0 + \rho\delta]$, or equivalently for $t \in [t_0, (t_0 + \delta) + W_{t_0+\delta}]$:

$$\forall n: \quad D_t(n) - D_{t_0}(n) = \frac{\lambda_1}{\rho^n}(t - t_0),$$

which means that for $t \in]t_0, t_0 + \delta[$, the queue of a homogeneous layer of composition $[\frac{\lambda_1}{\rho}, \frac{\lambda_1}{\rho^2}, \ldots, \frac{\lambda_1}{\rho^N}]$.

□

Remark 2.12
The condition: $\sum_{n=1}^{N-1} \frac{n\lambda_1}{\mu_{n+1}} < 1$ *may not be optimal, but it is easy to check and it will be satisfied in Bramson networks for the traffic intensities that we will choose.*

We are now going to use these results to exhibit diverging trajectories for a class of fluid networks.

3 Fluid Bramson networks.

3.1 Description of the model. Statement of the results.

The model that we are now going to study was presented by Bramson in [2]. It consists in two queues ($K = 2$) and one type of customers ($I = 1$); they enter the network at queue 1 at rate $\nu_1 = 1$ and leave it at the same queue, but meanwhile they visit queue 2 where they achieve $J \geq 1$ feed-backs; more precisely, $l_1 = J + 3$ and:

$$k_{11} = 1, k_{1s} = 2, s \leq 2 \leq J + 2, k_{1(J+3)} = 1.$$

We will maintain the index denoting the (single) type of customers in order to avoid a possible confusion between notations referring to classes (or equivalently to stages in the route) or queues.

At first we will prove the following, easy result.

Proposition 3.1
If $\sum_{s=1}^{J+3} \rho_{1s} < 1$, then there exists a time $T > 0$ such that:

$$\forall t \geq T, \forall s, 1 \leq s \leq J+3 : \quad Q_t(1,s) = 0,$$

for any fluid limit model of the Bramson network with J feed-backs.

Proof :
Let us denote:

$$Q(1,s) \triangleq \sum_{r \leq s} Q(1,r), \quad 1 \leq s \leq J+3,$$

and:

$$g(t) \triangleq \sum_{s=1}^{J+3} \frac{Q_t(1,s)}{\mu_{1s}}.$$

It is obvious that at any time time t where the derivatives exist, we have:

$$\begin{aligned} \dot{g}(t) &= \sum_{s=1}^{J+3} (\rho_{1s} - \frac{\dot{D}_t(1,s)}{\mu_{1s}}) \\ &= \sum_{s=1}^{J+3} \rho_{1s} - \dot{B}_t(1) - \dot{B}_t(2). \end{aligned}$$

If $g(t) > 0$, then $W_t(k) > 0$ for some queue $k = 1$ or 2, and then (by Lemma 2.6): $\dot{B}_t(k) = 1$, and in consequence:

$$\dot{g}(t) \leq \sum_{s=1}^{J+3} \rho_{1s} - 1 < 0.$$

The conclusion is then easy (see [4], Proposition 2.1 and Lemma 2.2 for a rigorous statement). □

The following corollary is a direct consequence of Theorem 2.5.

Corollary 3.2
If $\sum_{s=1}^{J+3} \rho_{1s} < 1$, then the Bramson network with J feed-backs is ergodic.

Remark 3.3
These results may obviously be generalized to any multiclass network.

In the following paragraph, we will study *exact* trajectories of the fluid model for the configuration:

$$\rho_{12} = \rho_{1(J+3)} = c, \quad \rho_{11} = \rho_{1s} = 0, \quad 3 \leq s \leq J+2. \tag{15}$$

Notice that since we set: $\nu_1 = 1$, traffic intensities and mean service times of their respective classes are equal.

A generic state of this network is pictured in figure 4. (it is easy to identify the route by following the arrows; only the external, arrival rate is written in the corresponding case). However, we have to face a delicate problem: the fluid limit models of a multiclass network are only defined for positive, mean service times (because the existence of limits of all the scaled processes is deduced from the convergence

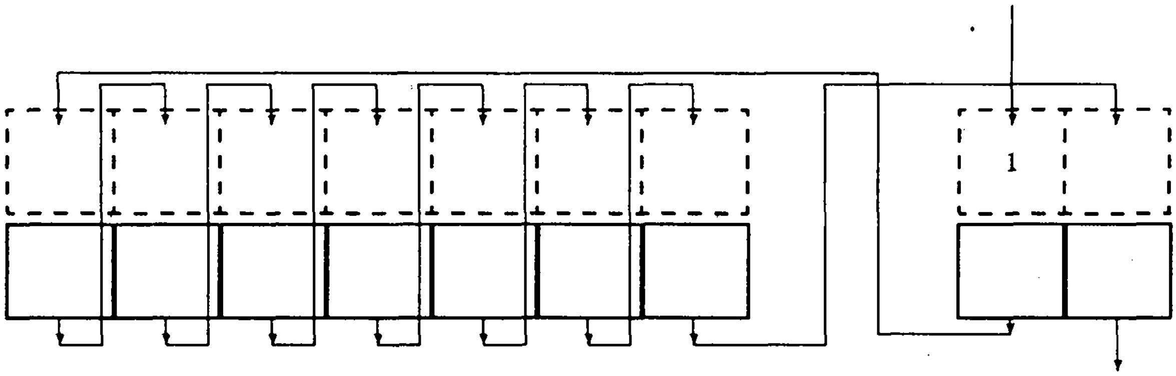

Figure 4 : Basic picture.

of $\overline{B}^n$ through relation (4)). Nevertheless, we may consider the limit paths that will be obtained by letting some mean service times tend to 0. The rules (given in paragraph 2.2) that govern the behaviour of the fluid model will be used for the configuration (15) in this spirit.

We can now state our main result. From now on, the fluid model is assumed to be in the configuration (15). Consider a fluid state of the network with the following characterics: queue 1 is empty, and queue 2 is a non-empty, homogeneous layer of composition:

$$[a_{1s}]_{2\leq s\leq J+2} = [c^J, c^{J-1}, ..., 1].$$

This kind of state will be called a "reference state" of the network. This definition is not arbitrary: it may be checked that if you concentrate all the initial customers at stage $(1, 2)$ (as Bramson did), then at the time when they will have experienced J successive services, the network will be in a reference state (a generalized version of this result is proven in Lemma 3.9). Thus we may consider a fluid limit model of a Bramson network starting from a reference state.

Theorem 3.4
For $J \geq 3$, the fluid Bramson network with J feed-backs in configuration (15) exhibits unstable cycles for $c \in]u_J, v_J[$, where $]u_J, v_J[$ is a non-empty interval included in $]0,1[$. More precisely, it is possible to define two sequences $(u_J)_{J\in\mathbb{N}}$ and $(v_J)_{J\in\mathbb{N}}$ that have the following properties:
(i) $(u_J)_{J\in\mathbb{N}}$ (resp. $(v_J)_{J\in\mathbb{N}}$) is a non-increasing (resp. a non-decreasing) sequence converging to $1/2$ (resp. to 1); for $J \geq 3$, we have: $\frac{1}{2} < u_J < v_J < 1$, and to each $c \in]u_J, v_J[$ are associated some characteristic constants $\theta(c) > 1$ and $T(c) > 0$;
(ii) if $J \geq 3$ and $c \in]u_J, v_J[$, the fluid model starting from a reference state follows a deterministic trajectory; if initially we have: $Q(1, J+2) = q$, then at time $t = T(c)q$ the fluid model returns to a reference state, and $Q(1, J+2) = \theta(c)q$; the same pattern is then repeated infinitely often.

In view of this result, we may conjecture that the stochastic network in configuration (15) is transient for $J \geq 3$ and $c \in]u_J, v_J[$. Unfortunately, there is no general result (for example a kind of large deviation result) supporting this idea. We will make some comments about these problem and related questions at the end of this paper.

Remark 3.5
Notice that as far as the fluid model in configuration (15) is concerned, Theorem 3.4 combined with Proposition 3.1 gives us a complete, asymptotic result when $J \to +\infty$: the network empties from any initial condition if $c \in [0, 1/2[$, and for $c \in]1/2, 1[$ there is a diverging trajectory which is the unique path starting from a reference state and satisfying the fluid equations.

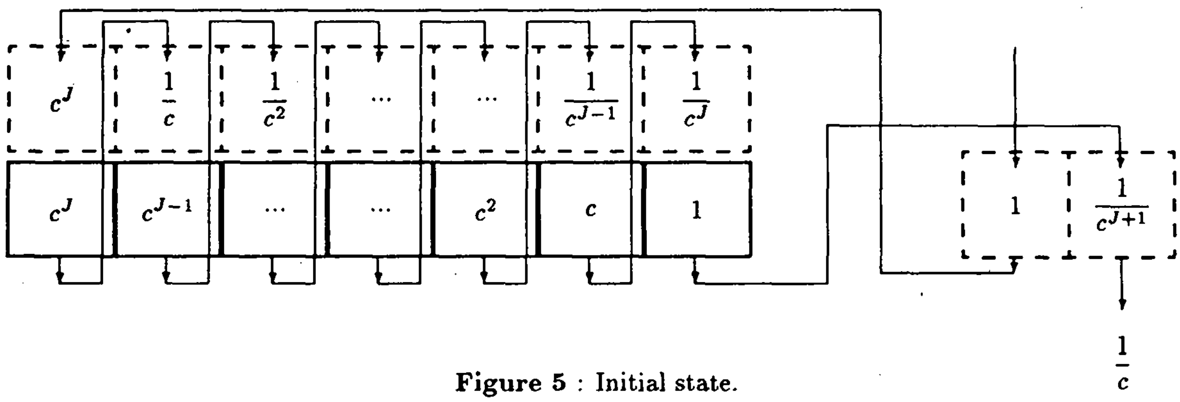

Figure 5 : Initial state.

The last section is devoted to the proof of Theorem 3.4. We emphasize the fact that we will not exhibit a special trajectory and check that it satisfies the fluid equations; these fluid equations will impose the form of our unstable trajectory.

3.2 Proof of Theorem 3.4.

The initial state is the state reached at time S_J in Bramson's proof. Before we start describing the evolution of the network from this initial state, notice that if there is a positive load in queue 2 at time t (i.e. $W_t(2) = cQ_t(1,2) > 0$), and the layer $]t + w_1, t + w_1 + w_2] \subseteq]t, t + W_t(2)]$ is a homogeneous layer of composition $a = [a_{12}, a_{13}, ..., a_{1(J+2)}]$, then: $|a/\mu| = a_{12}c$ (since all the classes but $(1,2)$ have null service times), and in consequence the departure rates for class $(1,s)$ customers in this layer is: $\dfrac{a_{1s}}{a_{12}c}$ (and in consequence class $(1,2)$ customers will leave queue 2 at rate $1/c$ regardless of the specific composition of the layer).

Now consider the situation at time $T_0 = 0$ as described by Figure 5. For $t \in]0, W_0(2)[$, we have:

$$\dot{D}_t(1,s) = \frac{a_{1s}}{a_{12}c} = \frac{c^{J-(s-2)}}{c^J c} = \frac{1}{c^{s-1}}, \quad 2 \le s \le J+2.$$

As a first consequence: $\dot{A}_t(1, J+3) = \dfrac{1}{c^{J+1}}$, and since $\dot{A}_t(1,1) = 1$ for all t, we may apply the results of Proposition 2.9 to queue 1: since the traffic intensity on $]0, W_0(2)[$ is: $\rho = \frac{c}{c^{J+1}} = \frac{1}{c^J} > 1$, the arrivals create a homogeneous layer of composition: $[a_{11}, a_{1(J+3)}] = [1, \frac{1}{c^{J+1}}]$, or equivalently $[c^{J+1}, 1]$; in consequence, for $t \in]0, W_0(2)[$, we have:

$$\dot{D}_t(1,1) = c^J, \quad \dot{D}_t(1, J+3) = \frac{1}{c}.$$

As for queue 2, we then have:

$$\dot{A}_t(1,2) = c^J, \quad \dot{A}_t(1,s) = \frac{1}{c^{s-2}}, \quad 3 \le s \le J+2,$$

which creates at the top of the queue a homogeneous layer of composition $[c^J, \frac{1}{c}, \frac{1}{c^2}, ..., \frac{1}{c^J}]$, or:

$$[c^{2J}, c^{J-1}, c^{J-2}, ..., 1].$$

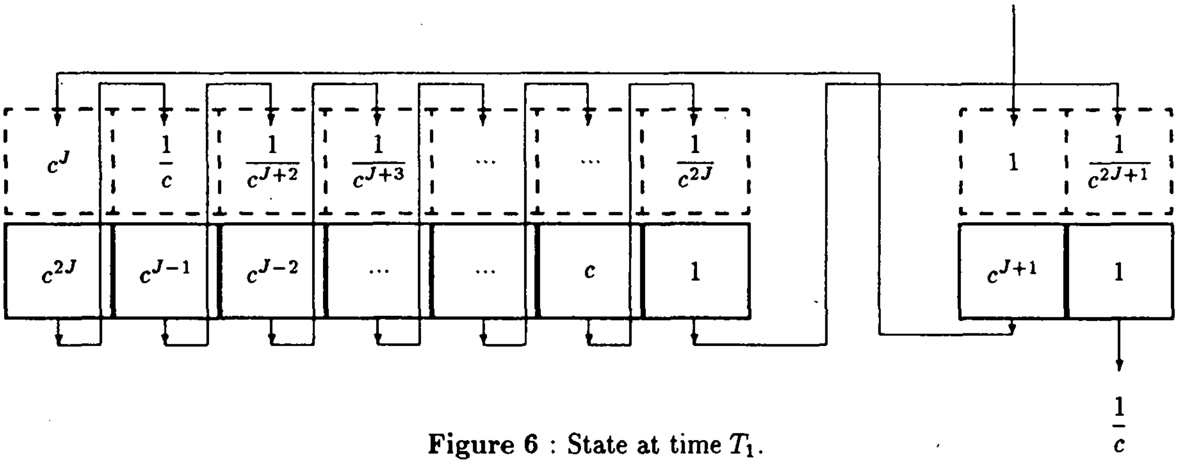

Figure 6 : State at time T_1.

Let $T_1 = W_0(2) = Q_0(1,2)c = c^{J+1}Q_0(1,J+2) = c^{J+1}q$. At this time the state of the network is characterized by a homogeneous layer of composition $[c^{J+1},1]$ at queue 1 and a homogeneous layer of composition $[c^{2J},c^{J-1},c^{J-2},...,1]$ at queue 2 (see Figure 6). To be complete, we must specify the thicknesses of these layers, or equivalently the "volume" of some class in each queue; for example, we have:

$$\begin{cases} Q_{T_1}(1,J+2) = \dfrac{1}{c^J}T_1 = cq \quad (= Q_0(1,J+1)) \\ Q_{T_1}(1,J+3) = (\dfrac{1}{c^{J+1}} - \dfrac{1}{c})T_1 = (1-c^J)q \end{cases}$$

Now the evolution of the state depends on which layer empties first, or which layer is less thick than the other. The answer obviously depends on the value of c. Since we want to find trajectories exhibiting the same kind of unstable behaviour as in Bramson's case, we will make an assumption reflecting the idea that queue 2 empties "quickly" from our initial state (which is the state reached at time S_J for Bramson); this assumption will later be translated into a constraint on c.

Assumption (H1): we assume that queue 2 empties before the homogeneous layer in queue 1 at time T_1 finishes being processed (that is before time $T_1 + W_{T_1}(1)$).

Under this assumption, we will be able to describe the behaviour of the network until the time at which queue 2 empties; then we will have to write that the residual volume of the original, homogeneous layer of queue 1 at this time is positive; our assumption (H1) will thus be translated into a constraint on c.

Let us consider the following sequence:

$$\begin{cases} T_0 = 0 \\ T_{j+1} = T_j + W_{T_j}(2) \end{cases}$$

We recall that for $j \geq 0$, $W_{T_j}(2)$ is the time needed to process all the customers waiting in queue 2 at time T_j. The evolution of the network on $[T_1, T_J]$ is described by the following lemma.

Lemma 3.6
Under assumption (H1), at time T_j $(1 \leq j \leq J)$ the network is in the following state (see Figure 7). At

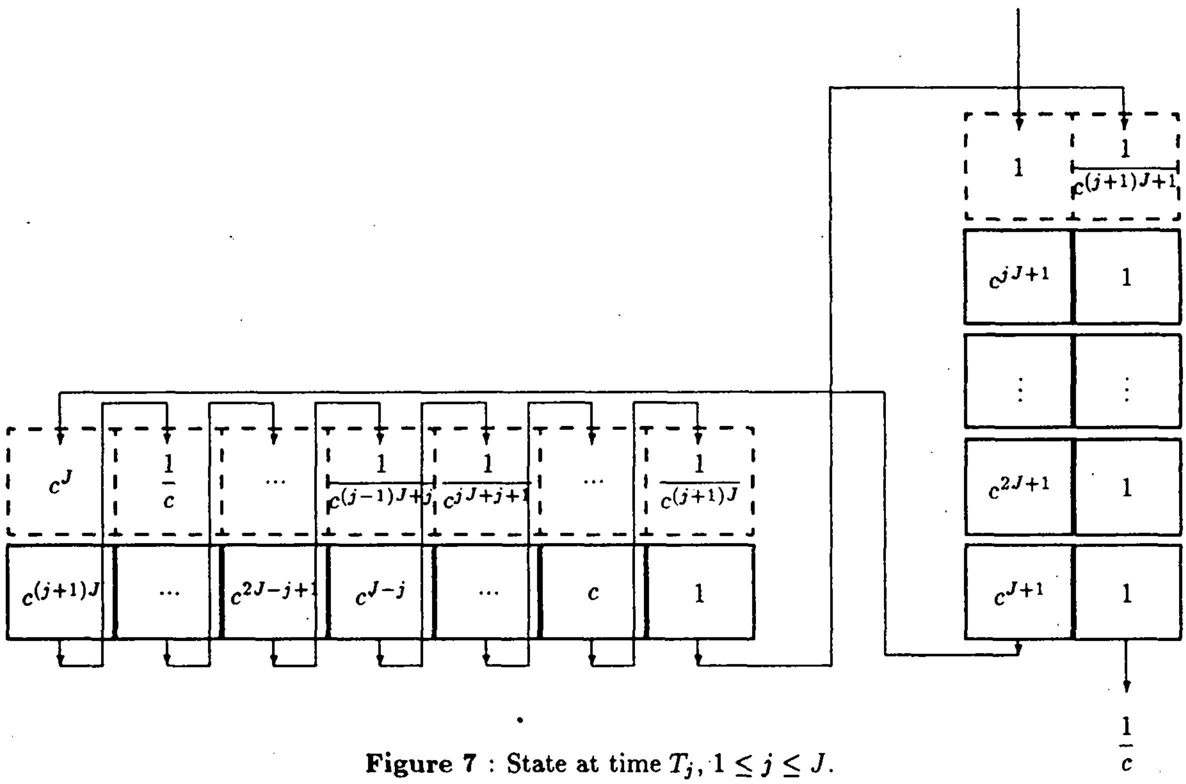

Figure 7 : State at time $T_j, 1 \le j \le J$.

queue 1, *there is a superposition of* j *homogeneous layers of compositions:* $[c^{iJ+1}, 1]$, $1 \le i \le j$. *Their thicknesses is determined by:*

$$Q^{(1)}_{T_j}(1, J+3) = (1 - \sum_{n=1}^{j} c^{n(J+1)-1})q, \quad Q^{(i)}_{T_j}(1, J+3) = c^{i-1}q, 1 < i \le j \ (if\ j > 1),$$

where $Q^{(i)}_{T_j}(1, J+3)$ *denotes the amount of class* $(1, J+3)$ *customers in the* i^{th} *homogeneous layer. At queue* 2 *there is a homogeneous layer of composition* $a = [a_{1s}]_{2 \le s \le J+2}$ *with:*

$$a_{1s} = c^{(j+1)J-(s-2)(J+1)}, 2 \le s \le j+2, \quad a_{1s} = c^{J-(s-2)}, j+2 \le s \le J+2.$$

Its thickness is determined by:

$$Q_{T_j}(1, J+2) = c^j q.$$

Proof :
We will prove this result by induction on j. It is already satisfied for $j = 1$. Assume that it is valid until time T_j with $1 \le j < J$ (and then $J > 1$). Then under assumption (H1), the departure rates for $t \in]T_j, T_{j+1}[$ are as given by Figure 7. More precisely:

$$\dot{D}_t(1,1) = c^J, \quad \dot{D}_t(1, J+3) = \frac{1}{c}, \text{ and:}$$

$$\dot{D}_t(1,s) = \frac{a_{1s}}{a_{12}c} = \frac{c^{(j+1)J-(s-2)(J+1)}}{c^{(j+1)J}c} = \frac{1}{c^{(s-2)(J+1)+1}}, \quad 2 \le s \le j+2,$$
$$= \frac{c^{J-(s-2)}}{c^{(j+1)J}c} = \frac{1}{c^{jJ+s-1}}, \quad j+2 \le s \le J+2.$$

Moreover, we have:

$$T_{j+1} - T_j = W_{T_j}(2) = Q_{T_j}(1,2)c = c^{(j+1)J}Q_{T_j}(1, J+2)c = c^{(j+1)J}c^j cq = c^{(j+1)(J+1)}q.$$

At time T_{j+1} in queue 1, under assumption (H1), the original, homogeneous layer of time T_1 has not disappeared yet, and then the i^{th} homogeneous layer of time T_j for $1 < i \leq j$ (if $j > 1$) are preserved. The residual thickness of the first, homogeneous layer is determined by:

$$\begin{aligned} Q^{(1)}_{T_{j+1}}(1, J+3) &= Q^{(1)}_{T_j}(1, J+3) - \int_{T_j}^{T_{j+1}} \dot{D}_t(1, J+3)dt \\ &= (1 - \sum_{n=1}^{j} c^{n(J+1)-1})q - \frac{1}{c}(T_{j+1} - T_j) \\ &= (1 - \sum_{n=1}^{j+1} c^{n(J+1)-1})q. \end{aligned}$$

Then we may apply Proposition 2.9 again. At queue 1, the arrival rates are:

$$\dot{A}_t(1,1) = 1, \quad \dot{A}_t(1, J+3) = \dot{D}_t(1, J+2) = \frac{1}{c^{(j+1)J+1}},$$

which generates at the top of the queue a $(j+1)^{th}$ homogeneous layer of composition $[c^{(j+1)J+1}, 1]$. Under assumption (H1) again, at time T_{j+1} its thickness is determined by:

$$\begin{aligned} Q^{(j+1)}_{T_{j+1}}(1, J+3) &= \int_{T_j}^{T_{j+1}} \dot{A}_t(1, J+3)dt \\ &= D_{T_{j+1}}(1, J+2) - D_{T_j}(1, J+2) \\ &= D_{T_j + W_{T_j}(2)}(1, J+2) - D_{T_j}(1, J+2) \\ &= A_{T_j}(1, J+2) - D_{T_j}(1, J+2) \\ &= Q_{T_j}(1, J+2) = c^j q. \end{aligned}$$

As for queue 2, the arrival rates are:

$$\begin{aligned} \dot{A}_t(1,2) = c^J, \quad \dot{A}_t(1,s) &= \frac{1}{c^{(s-3)(J+1)+1}}, \quad 3 \leq s \leq j+3, \\ &= \frac{1}{c^{jJ+s-2}}, \quad j+3 \leq s \leq J+2, \end{aligned}$$

so that a homogeneous layer is generated according to the composition given in the lemma (which is obtained by multiplying the arrival rates by $c^{(j+1)J}$). The layer of time T_j disappears at time T_{j+1}, and at this time we easily check that:

$$Q_{T_{j+1}}(1, J+2) = Q_{T_j}(1, J+1) = cQ_{T_j}(1, J+2) = c^{j+1}q.$$

The proof is now complete.

□

Let us examine the state of the network at time T_J (see Figure 8). At queue 1, there is a superposition of J homogeneous layers of compositions: $[c^{iJ+1}, 1]$, $1 \leq i \leq J$. Their thicknesses is determined by:

$$Q^{(1)}_{T_J}(1, J+3) = (1 - \sum_{n=1}^{J} c^{n(J+1)-1})q, \quad Q^{(i)}_{T_J}(J+3) = c^{i-1}q, 1 < i \leq J \text{ (if } J > 1).$$

At queue 2 there is a homogeneous layer of composition $a = [a_{1s}]_{2 \leq s \leq J+2}$ with:

$$a_{1s} = c^{(J+1)J-(s-2)(J+1)} = c^{(J+1)(J+2-s)}, 2 \leq s \leq J+2.$$

Its thickness is determined by:

$$Q_{T_J}(1, J+2) = c^J q.$$

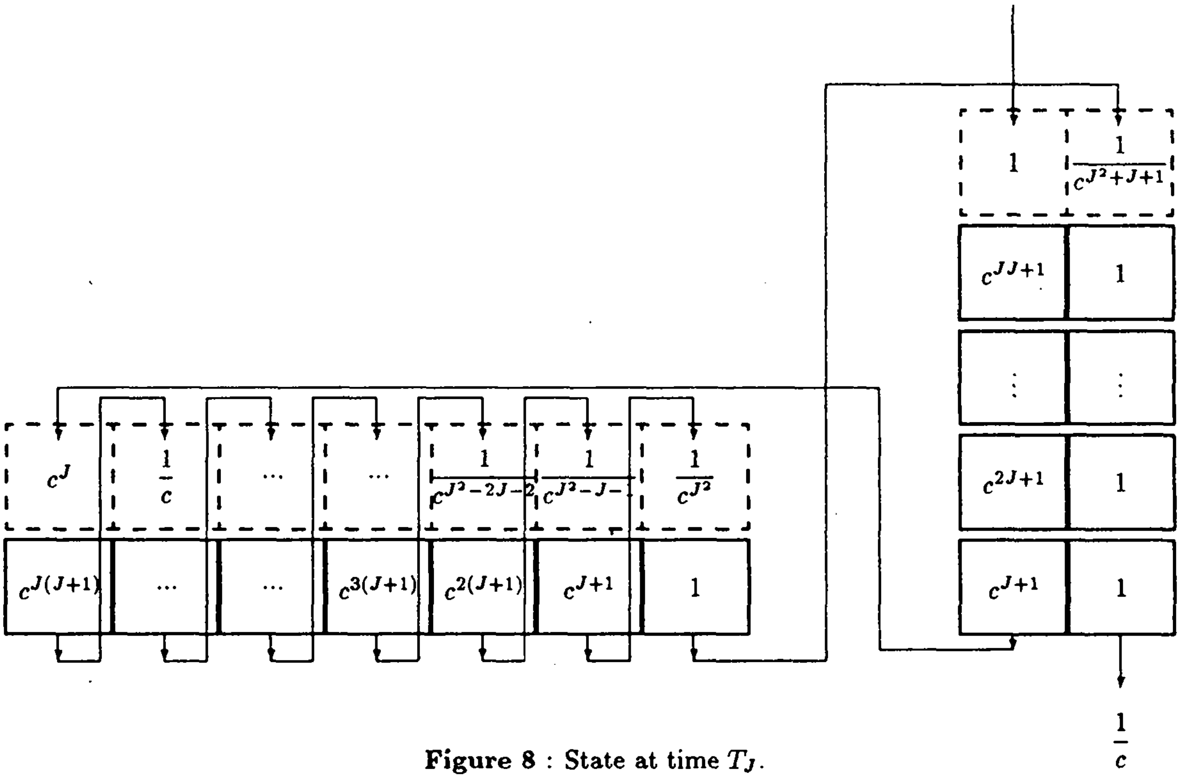

Figure 8 : State at time T_J.

For $t \in]T_J, T_{J+1}[$, the arrival rates in queue 2 are (under assumption (H1)):

$$\dot{A}_t(1,2) = c^J, \quad \dot{A}_t(1,s) = \frac{1}{c^{(s-3)(J+1)+1}}, \quad 3 \le s \le J+2, \quad \text{or:}$$

$$\dot{A}_t(1,s) = \frac{c^{(J+1)(J+2-s)}}{c^{J^2}}, \quad 2 \le s \le J+2.$$

This means that the new homogeneous layer generated at the top of queue 2 has the same composition as the previous one.

In view of assumption (H1), this implies that the flows of fluids (or the departure rates) through the network will not change any more until queue 2 empties, which will occur at some time T satisfying:

$$\begin{aligned} Q_T(1,J+2) &= Q_{T_J}(1,J+2) + \int_{T_J}^{T} (\dot{A}_t(1,J+2) - \dot{D}_t(1,J+2))dt \\ &= c^J q + (\frac{1}{c^{J^2}} - \frac{1}{c^{J^2+J+1}})(T - T_J) \\ &= 0, \end{aligned}$$

or:

$$T - T_J = \frac{c^J q}{\frac{1}{c^{J^2+J+1}} - \frac{1}{c^{J^2}}} = \frac{c^{(J+1)^2}}{1 - c^{J+1}} q.$$

At time T, queue 2 is empty, and at queue 1 there is a superposition of $J+1$ homogeneous layers of compositions: $[c^{iJ+1}, 1]$, $1 \le i \le J+1$ (see Figure 9). We have:

$$Q_T^{(i)}(1, J+3) = c^{i-1} q, 1 < i \le J \text{ (if } J > 1\text{), and:}$$

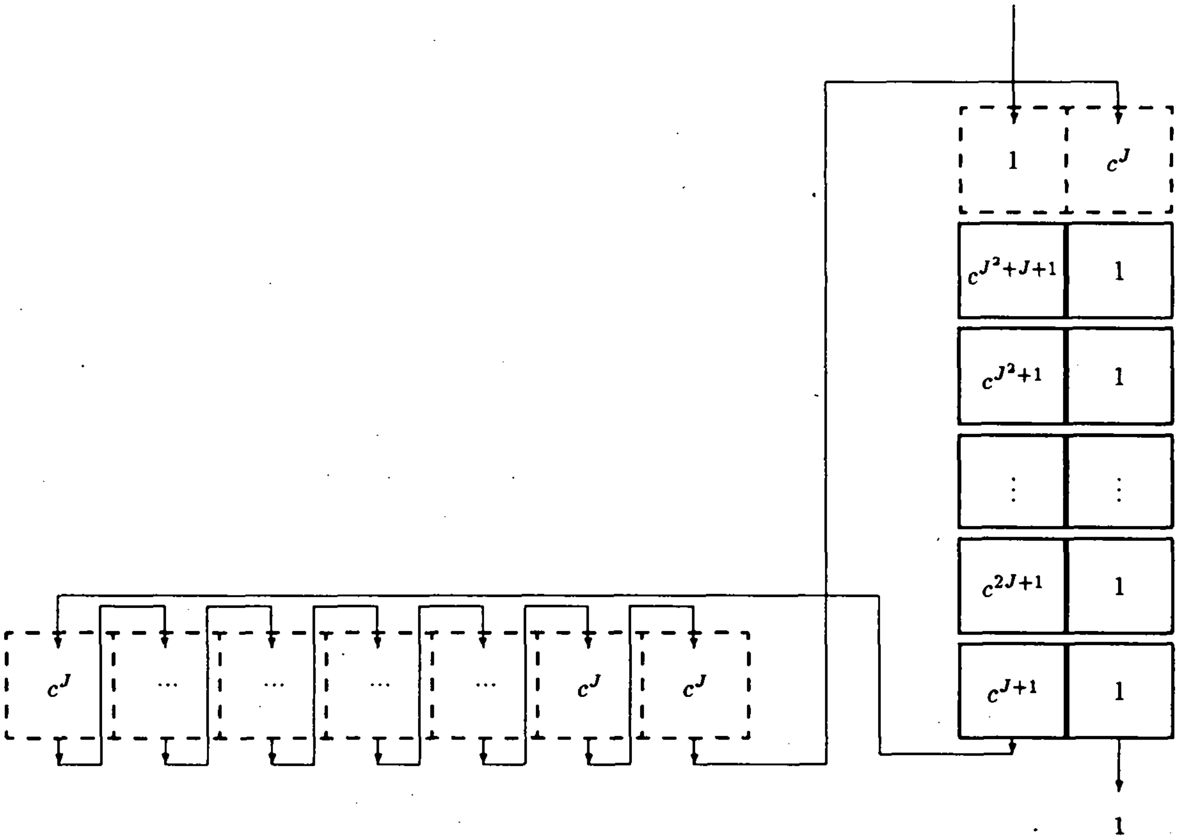

Figure 9 : State at time T.

$$\begin{aligned} Q_T^{(J+1)}(1, J+3) &= \int_{T_J}^{T} \dot{D}_t(1, J+3)dt \\ &= \frac{1}{c^{J^2+J+1}}(T - T_J) \\ &= \frac{1}{c^{J^2+J+1}} \frac{c^{(J+1)^2}}{1 - c^{J+1}} q \\ &= \frac{c^J}{1 - c^{J+1}} q \quad \text{and:} \end{aligned}$$

$$\begin{aligned} Q_T^{(1)}(1, J+3) &= Q_{T_J}^{(1)}(1, J+3) - \frac{1}{c}(T - T_J) \\ &= (1 - \sum_{n=1}^{J} c^{n(J+1)-1})q - \frac{c^{J(J+2)}}{1 - c^{J+1}} q \\ &= \frac{1 - c^J - c^{J+1}}{1 - c^{J+1}}. \end{aligned}$$

In consequence, assumption (H1) amounts to:

Condition (H1): c must be chosen such that $Q_T^{(1)}(1, J+3) > 0$, or: $\boxed{1 - c^J - c^{J+1} > 0}$.

Let us now examine the situation at time: $U_0 = T + W_T(1)$, when all the customers waiting in queue 1 at time T will have been served.

Lemma 3.7
Queue 2 remains empty during the period $[T, U_0]$. At time U_0 in queue 1, there is a superposition of $J + 1$

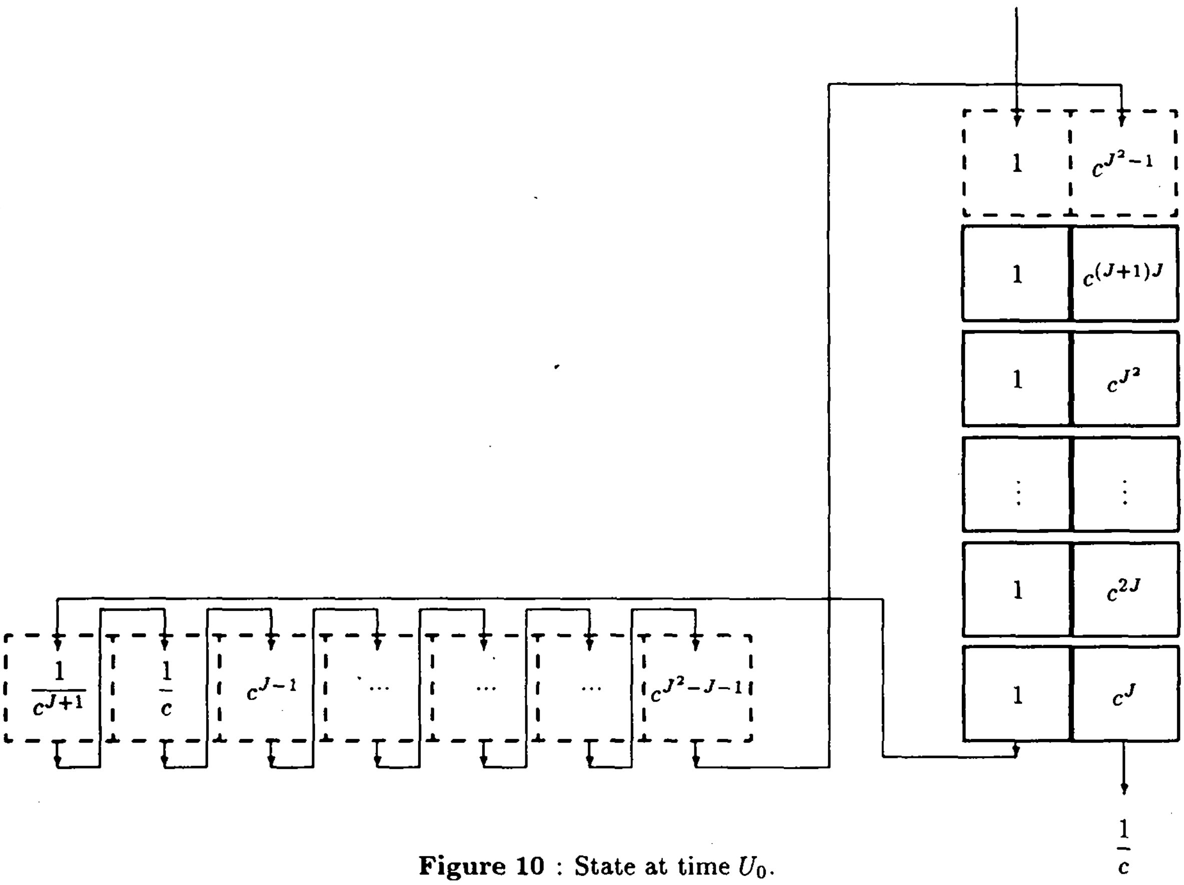

Figure 10 : State at time U_0.

homogeneous layers of compositions: $[1, c^{iJ}]$, $1 \le i \le J+1$ *(see Figure 10), and their thicknesses is given by:*

$$\begin{cases} Q_{U_0}^{(1)}(1,1) = c\dfrac{1 - c^J - c^{J+1}}{1 - c^{J+1}}q, \\ Q_{U_0}^{(i)}(1,1) = c^i q, \quad 1 < i \le J \ (\textit{if } J > 1), \\ Q_{U_0}^{(J+1)}(1,1) = \dfrac{c^{J+1}}{1 - c^{J+1}}q, \end{cases}$$

where $Q_{U_0}^{(i)}(1,1)$ *denotes the amount of class* $(1,1)$ *customers in the* i^{th} *homogeneous layer.*

Proof :
According to the description of the state of queue 1 at time T, we have, for $t \in]T, T + W_T(1)[$:

$$\begin{aligned} \dot{A}_t(1,2) = \dot{D}_t(1,1) \quad &= c^{iJ} \quad \text{if the } i^{th} \text{ layer is being processed} \\ &\le c^J \quad \text{in any case,} \end{aligned}$$

which implies that the potential traffic intensity is always less than 1. Then a simple application of Proposition 2.11 shows that queue 2 remains empty on $[T, U_0]$. In consequence, at any time $t \in]T, U_0[$ where the derivative exists, we have:

$$\dot{A}_t(1, J+3) = \dot{D}_t(1,1), \quad \text{whereas: } \dot{A}_t(1,1) = 1.$$

It is then easy to see (in view of Proposition 2.9) that the homogeneous layer of composition $[1, c^{iJ+1}]$ in queue 1 at time T generates a homogeneous layer of composition $[c^{iJ}, 1]$ in queue 1 at time U_0. Moreover, the thickness of the i^{th} homogeneous layer at time U_0 is determined by:

$$Q_{U_0}^{(i)}(1, J+3) = Q_T^{(i)}(1,1), \quad \text{or equivalently:}$$

$$Q_{U_0}^{(i)}(1,1) = cQ_T^{(i)}(1, J+3),$$

which completes our proof. □

Thus at time U_0, queue 2 is empty and in queue 1 there is a superposition of $(J+1)$ homogeneous layers; their thicknesses will be denoted by: $w^{(i)}$, $1 \le i \le J+1$ (which satisfy: $\sum_{i=1}^{J+1} w^{(i)} = W_{U_0}(1)$).

Consider time $U_1 = U_0 + w^{(1)}$. For $t \in]U_0, U_1[$:

$$\dot{A}_t(1,2) = \dot{D}_t(1,1) = \frac{1}{c^{J+1}},$$

so that the potential traffic intensity for queue 2 on this time interval is $\frac{1}{c^J} > 1$. We may apply Proposition 2.11 again. Using the notations introduced in this proposition, we have: $f(x) = \frac{1}{c^J}$, and then: $\dot{f}(1) = 0 > -1$ and: $\rho = \frac{1}{c^J}$. In consequence, for $t \in]U_0, U_1[$, queue 2 is a homogeneous layer of composition:

$$[a_{1s}]_{2 \le s \le J+2} = [\frac{1}{c^{J+1}} c^{(s-1)J}]_{2 \le s \le J+2},$$

which means that:

$$\dot{D}_t(1,s) = c^{(s-2)J-1}, \quad 2 \le s \le J+2.$$

At queue 1, we thus have the following arrival rates:

$$\dot{A}_t(1,1) = 1, \quad \dot{A}_t(1,J+3) = c^{J^2-1},$$

which generates a new, homogeneous layer of composition $[1, c^{J^2-1}]$.

Then the state of the network at time U_1 may be described as follows (see Figure 11). At queue 1, there is a superposition of $(J+1)$ homogeneous layers. The i^{th} layer, for $1 \le i \le J$, has composition $[1, c^{(i+1)J}]$ and $Q_{U_1}^{(i)}(1,1) = c^{i+1}q$. The $(J+1)^{th}$ layer has composition $[1, c^{J^2-1}]$ and:

$$\begin{aligned} Q_{U_1}^{(J+1)}(1,1) &= \int_{U_0}^{U_1} \dot{A}_t(1,1)dt = U_1 - U_0 = w^{(1)} \\ &= Q_{U_0}^{(1)}(1,J+3)c = c^J Q_{U_0}^{(1)}(1,1)c \\ &= c^{J+2}\frac{1-c^J-c^{J+1}}{1-c^{J+1}}q. \end{aligned} \tag{16}$$

At queue 2 there is a homogeneous layer of composition $[1, c^J, c^{2J}, ..., c^{J^2}]$ and:

$$\begin{aligned} Q_{U_1}(1,2) &= Q_{U_0}(1,2) + \int_{U_0}^{U_1} (\dot{A}_t(1,2) - \dot{D}_t(1,2))dt = (\frac{1}{c^{J+1}} - \frac{1}{c})(U_1 - U_0) \\ &= c(1-c^J)\frac{1-c^J-c^{J+1}}{1-c^{J+1}}q. \end{aligned} \tag{17}$$

Now we are in a similar situation as at time T_1: the future evolution of the state depends on which layer empties first, which again depends on the value of c. We will make a new assumption reflecting the idea that queue 1 empties "quickly", and this assumption will later be translated into a constraint on c.

Assumption (H2): we assume that queue 1 empties before the homogeneous layer in queue 2 at time U_1 finishes being processed (that is before time $U_1 + W_{U_1}(2)$).

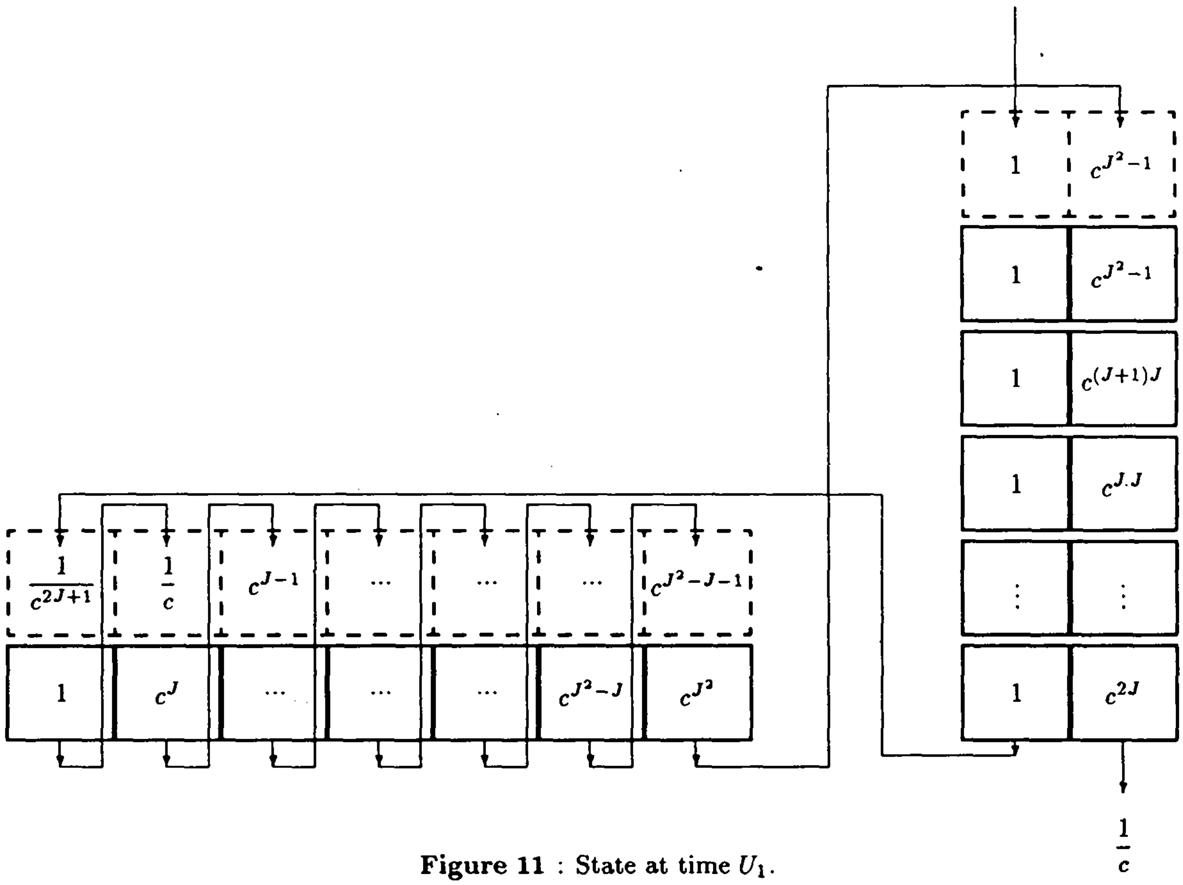

Figure 11 : State at time U_1.

Thus we may consider the following sequence:

$$\begin{cases} U_0 = T + W_T(1), \\ U_{j+1} = U_j + w^{(j+1)}, & \text{for } j \le J. \end{cases}$$

The state of the network at time $U_{J+1} = U_0 + W_{U_0}(1)$ is described by the following lemma.

Lemma 3.8
Under assumption (H2), at time U_{J+1}, the network is in the following state (see Figure 12). At queue 1, there is a homogeneous layer of composition $[1, c^{J^2-1}]$ and:

$$Q_{U_{J+1}}(1,1) = c^{J+2}\frac{1-c^J}{1-c^{J+1}}.$$

At queue 2, there is a superposition of $J+1$ homogeneous layers of respective compositions $a^{(i)} = [a^{(i)}_{1s}]_{2\le s\le J+2}$, $1 \le i \le J+1$, with:

$$a^{(i)}_{12} = 1, \quad a^{(i)}_{1s} = c^{(i+s-3)J}, \quad 2 < i \le J+2.$$

For $1 < i < J+1$ (if $J > 1$), we have: $Q^{(i)}_{U_{J+1}}(1,2) = c^i q$, whereas:

$$Q^{(1)}_{U_{J+1}}(1,2) = c\frac{(1-c^J)^2 - c^{J+1}}{1-c^{J+1}}, \quad Q^{(J+1)}_{U_{J+1}}(1,2) = \frac{c^{J+1}}{1-c^{J+1}}.$$

Proof :

At first, let us calculate $U_{J+1} - U_0$, that is $W_{U_0}(1)$:

$$\begin{aligned}
W_{U_0}(1) &= Q_{U_0}(1, J+3)c = c\sum_{i=1}^{J+1} Q_{U_0}^{(i)}(1, J+3) \\
&= c\sum_{i=1}^{J+1} c^{iJ} Q_{U_0}^{(i)}(1,1) \\
&= cq\left[c^J c\frac{1-c^J-c^{J+1}}{1-c^{J+1}} + \sum_{i=2}^{J} c^{iJ}c^i + c^{(J+1)J}\frac{c^{J+1}}{1-c^{J+1}}\right] \\
&= \frac{cq}{1-c^{J+1}}\left[c^{J+1}(1-c^J-c^{J+1}) + c^{2(J+1)}(1-c^{(J-1)(J+1)}) + c^{(J+1)^2}\right] \\
&= c^{J+2}\frac{1-c^J}{1-c^{J+1}}q.
\end{aligned}$$

Then, assumption (H2) ensures us that for $t \in]U_1, U_{J+1}[$, we have:

$$\dot{D}_t(1,s) = c^{(s-2)J-1}, \quad 2 \le s \le J+2,$$

that is the same rates as for $t \in]U_0, U_1[$. The first consequence is that we have:

$$\dot{A}_t(1,1) = 1, \quad \text{and: } \dot{A}_t(1, J+3) = \dot{D}_t(1, J+2) = c^{J^2-1},$$

for $t \in]U_0, U_{J+1}[$, and since $U_{J+1} = U_0 + W_{U_0}(1)$, at time U_{J+1} in queue 1 there is a homogeneous layer of composition $[1, c^{J^2-1}]$ (see Proposition 2.9) and:

$$\begin{aligned}
Q_{U_{J+1}}(1,1) &= Q_{U_0+W_{U_0}(1)}(1,1) = A_{U_0+W_{U_0}(1)}(1,1) - A_{U_0}(1,1) \\
&= W_{U_0}(1) = c^{J+2}\frac{1-c^J}{1-c^{J+1}}q. \qquad (18)
\end{aligned}$$

As for queue 2, we have, for $t \in]U_j, U_{j+1}[$, $1 \le j \le J$:

$$\begin{cases} \dot{A}_t(1,2) = \dot{D}_t(1,1) = \dfrac{1}{c^{(j+1)J+1}} \\ \dot{A}_t(1,s) = \dot{D}_t(1,s-1) = c^{(s-3)J-1}, \quad 2 < s \le J+2, \end{cases}$$

which generates a $(j+1)^{th}$ homogeneous layer of composition $[1, c^{(j+1)J}, c^{(j+2)J}, ..., c^{(j+J)J}]$. In view of assumption (H2) (there is no departure from these layers up to time U_{J+1}), it is easy to check that: $Q_{U_{J+1}}^{(j+1)}(1,2) = Q_{U_0}^{(j+1)}(1,1)$ (the $(j+1)^{th}$ layer in queue 1 at time U_0 generated the $(j+1)^{th}$ layer in queue 2 at time U_{J+1}).

Thus the last calculation consists in estimating the thickness at time U_{J+1} of the first homogeneous layer in queue 2. Obviously we have:

$$\begin{aligned}
Q_{U_{J+1}}^{(1)}(1,2) &= Q_{U_1}(1,2) - \int_{U_1}^{U_{J+1}} \dot{D}_t(1,2)dt = Q_{U_1}(1,2) - \frac{1}{c}(U_{J+1} - U_1) \\
&= Q_{U_1}(1,2) - \frac{1}{c}((U_{J+1} - U_0) - (U_1 - U_0)) = Q_{U_1}(1,2) - \frac{1}{c}(W_{U_0}(1) - w^{(1)}),
\end{aligned}$$

with:

$$\begin{cases} Q_{U_1}(1,2) = c(1-c^J)\dfrac{1-c^J-c^{J+1}}{1-c^{J+1}}q \quad \text{(see (17))} \\ W_{U_0}(1) = c^{J+2}\dfrac{1-c^J}{1-c^{J+1}}q \quad \text{(see (18))} \\ w^{(1)} = c^{J+2}\dfrac{1-c^J-c^{J+1}}{1-c^{J+1}}q \quad \text{(see (16))} \end{cases}$$

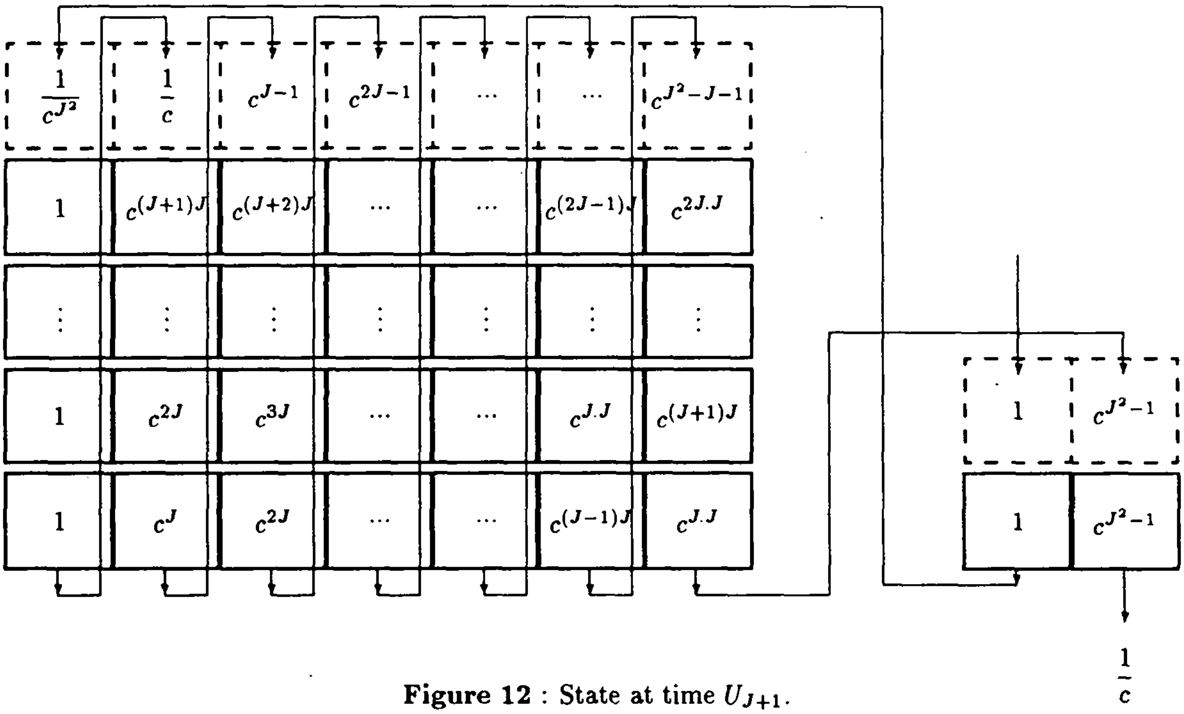

Figure 12 : State at time U_{J+1}.

After simplification we obtain:

$$Q^{(1)}_{U_{J+1}}(1,2) = c\frac{(1-c^J)^2 - c^{J+1}}{1-c^{J+1}}q. \tag{19}$$

The proof is now complete. □

Under assumption (H2), the arrival rates in queue 1 will be:

$$\dot{A}_t(1,1) = 1, \quad \dot{A}_t(1,J+3) = c^{J^2-1},$$

until queue 1 empties, which means that the state of queue 1 will be a homogeneous layer of composition $[1, c^{J^2-1}]$ from U_{J+1} until the time U where it will be empty. The flows of fluids (or the departure rates) through the network will not change any more until time U, which is then determined by:

$$\begin{aligned} Q_U(1,1) &= Q_{U_{J+1}}(1,1) + \int_{U_{J+1}}^{U} (\dot{A}_t(1,1) - \dot{D}_t(1,1))dt \\ &= c^{J+2}\frac{1-c^J}{1-c^{J+1}}q + (1-\frac{1}{c^{J^2}})(U - U_{J+1}) \quad \text{(see(18))} \\ &= 0, \end{aligned}$$

or:

$$U - U_{J+1} = \frac{c^{J^2}}{1-c^{J^2}}c^{J+2}\frac{1-c^J}{1-c^{J+1}}q.$$

At time U, queue 1 is empty, and at queue 2 there is a superposition of $J+2$ homogeneous layers (see Figure 13). The layers i, $1 < i < J+2$, are the same as at time U_{J+1} in view of assumption (H2).

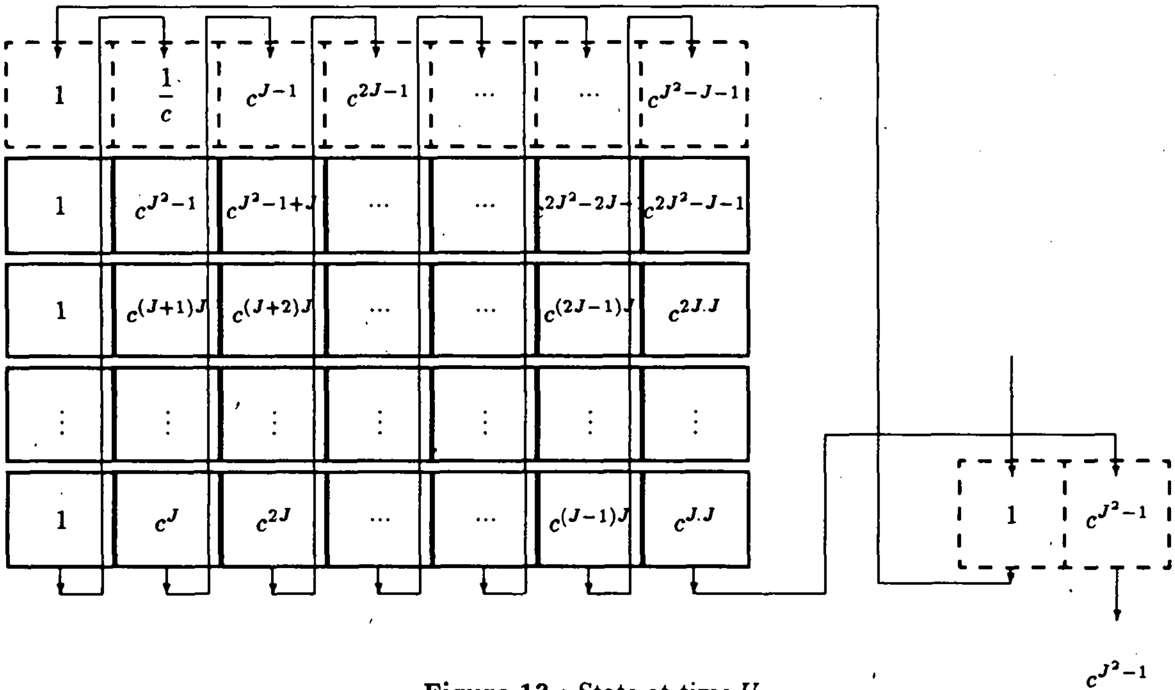

Figure 13 : State at time U.

The $(J+2)^{th}$ layer is generated by the arrivals on $]U_{J+1}, U[$, that is:

$$\begin{cases} \dot{A}_t(1,2) = \dot{D}_t(1,1) = \dfrac{1}{c^{J^2}} \\ \dot{A}_t(1,s) = \dot{D}_t(1,s-1) = c^{(s-3)J-1}, \quad 2 < s \leq J+2. \end{cases}$$

Thus at time U, we have a $(J+2)^{th}$ homogeneous layer of composition $a^{(J+2)} = [a^{(J+2)}_{1s}]_{2 \leq s \leq J+2}$, with:

$$a^{(J+2)}_{12} = 1, \quad a^{(J+2)}_{1s} = c^{J^2+(s-3)J-1}, \quad 2 < s \leq J+2.$$

Its thickness is determined by:

$$\begin{aligned} Q^{(J+2)}_U(1,2) &= A_U(1,2) - A_{U_{J+1}}(1,2) = \frac{1}{c^{J^2}}(U - U_{J+1}) \\ &= \frac{c^{J+2}(1-c^J)}{(1-c^{J^2})(1-c^{J+1})} q. \end{aligned}$$

The first layer still has composition $[1, c^J, c^{2J}, ..., c^{J^2}]$, but the volume of class $(1,2)$ customers in this layer is now:

$$\begin{aligned} Q^{(1)}_U(1,2) &= Q^{(1)}_{U_{J+1}}(1,2) - \int_{U_{J+1}}^{U} \dot{D}_t(1,2)dt \\ &= c\frac{(1-c^J)^2 - c^{J+1}}{1-c^{J+1}} q - \frac{1}{c}(U - U_{J+1}) \quad \text{(see (19))} \\ &= c\frac{(1-c^J)^2 - c^{J+1}}{1-c^{J+1}} q - \frac{1}{c}\frac{c^{J^2}}{1-c^{J^2}} c^{J+2} \frac{1-c^J}{1-c^{J+1}} q \quad \text{(see above)} \\ &= c\frac{(1-c^J-c^{J+1})(1-c^{J^2}) - c^J(1-c^J)}{(1-c^{J^2})(1-c^{J+1})} q. \end{aligned}$$

Now it is clear that assumption (H2) is equivalent to

Condition (H2): c must be chosen such that $Q_U^{(1)}(1,2) > 0$, or: $\boxed{1 - c^J - c^{J+1} > c^J \frac{1-c^J}{1-c^{J^2}}}$.

It is also clear that this condition is stronger than condition (H1).

The following lemma describes how the network comes back to a reference state (that is a state similar to the initial one).

Lemma 3.9
Assume that at some time V_0, queue 1 is empty, and in queue 2 there is a superposition of L homogeneous layers of respective compositions $a^{(l)} = [a_{12}^{(l)}, a_{13}^{(l)}, ..., a_{1(J+2)}^{(l)}]$, $1 \le l \le L$, with:

$$\forall l, 1 \le l \le L: \quad a_{12}^{(l)} = 1 \text{ and: } \quad \forall s, 2 < s \le J+2: \quad a_{1s}^{(l)} < c^{J-(s-2)}.$$

Denote: $Q_{V_0}(1,2) = q'$.

Consider the sequence: $V_{j+1} = V_j + W_{V_j}(2)$, $j \ge 0$. Then at time V_j, $0 \le j \le J$, the network is in the following state (see Figure 14). Queue 1 is empty, and at queue 2 there is a superposition of L homogeneous layers of respective compositions: $a^{(l)}(j) = [a_{12}^{(l)}(j), a_{13}^{(l)}(j), ..., a_{1(J+2)}^{(l)}(j)]$, $1 \le l \le L$, with:

$$\forall l, 1 \le l \le L: \quad a_{1s}^{(l)}(j) = c^{j-(s-2)}, 2 \le s \le j+2 \text{ and: } \quad a_{1s}^{(l)}(j) = a_{1(s-j)}^{(l)}, j+2 < s \le J+2 \text{ (if } j < J).$$

Moreover, we have: $Q_{V_j}(1, j+2) = q'$.

In consequence, at time V_J, queue 1 is empty and in queue 2 there is a homogeneous layer of composition $[c^J, c^{J-1}, c^{J-2}, ..., 1]$ (that is a reference state), and: $Q_{V_J}(1, J+2) = q'$.

Remark 3.10
This return to a reference state is not a miracle. Here the fact that all the classes but one have null services in queue 2 is essential, because as long as queue 1 remains empty, the class $(1,2)$ is fed by arrivals at rate 1 and is processed at rate $\frac{1}{c}$, which progressively regularizes the composition of the whole queue.

Proof :
We will prove this result by induction on j. It is valid for $j = 0$. Assume that it is true for some $j < J$. Then for $t \in]V_j, V_{j+1}[$, $2 \le s \le J+2$:

$$\begin{aligned}
\dot{D}_t(1,s) &= \frac{a_{1s}^{(l)}(j)}{a_{12}^{(l)}(j)c} \quad \text{if the } l^{th} \text{ layer is being processed} \\
&= \frac{c^{j-(s-2)}}{c^j c} = \frac{1}{c^{s-1}} \quad \text{if } 2 \le s \le j+2 \\
&= \frac{a_{1(s-j)}^{(l)}}{c^j c} \quad \text{if } j+2 < s \le J+2
\end{aligned}$$

Then at queue 1 we have:

$$\begin{aligned}
\dot{A}_t(1,1) = 1, \quad \dot{A}_t(1, J+3) &= \dot{D}_t(1, J+2) = \frac{a_{1(J+2-j)}^{(l)}}{c^j c} \quad \text{if the } l^{th} \text{ layer is being processed} \\
&< \frac{c^{J-((J+2-j)-2)}}{c^j c} = \frac{1}{c} \quad \text{by initial assumption.}
\end{aligned}$$

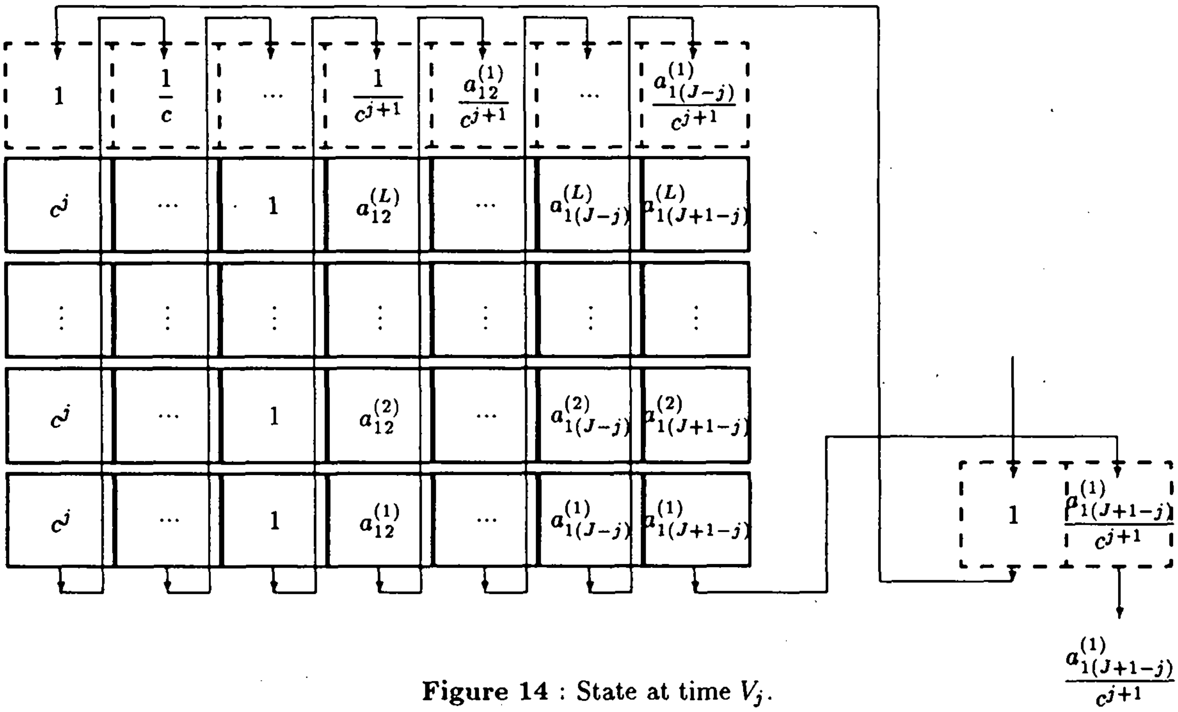

Figure 14 : State at time V_j.

In particular, this means that the traffic intensity at queue 1 is always less than 1, so that queue 1 remains empty.

Then, while the l^{th} homogeneous layer in queue 2 at time V_j is being processed, we have:

$$\begin{cases} \dot{A}_t(1,2) = \dot{D}_t(1,1) = \dot{A}_t(1,1) = 1 \quad \text{(queue 1 empty)}, \\ \dot{A}_t(1,s) \;\; = \dot{D}_t(1,s-1) = \dfrac{1}{c^{s-2}} \quad \text{if } 2 < s \leq j+3, \\ \qquad\qquad = \dfrac{a^{(l)}_{1(s-1-j)}}{c^j c} \quad \text{if } j+1 < J \text{ and } j+3 < s \leq J+2. \end{cases}$$

Hence, the treatment of the l^{th} homogeneous layer generates a new homogeneous layer at the top of queue 2; it is obvious that this new layer will be the l^{th} one at time V_{j+1}; its composition is: $a^{(l)}(j+1) = [a^{(l)}_{12}(j+1), a^{(l)}_{13}(j+1), ..., a^{(l)}_{1(J+2)}(j+1)]$, with:

$$a^{(l)}_{1s}(j+1) = \frac{1}{c^{s-2}}, 2 \leq s \leq j+3 \text{ and: } \; a^{(l)}_{1s}(j+1) = \frac{a^{(l)}_{1(s-1-j)}}{c^j c}, j+3 < s \leq J+2 \text{ (if } j+1 < J),$$

or equivalently (we may multiply everything by c^{j+1}):

$$\begin{cases} a^{(l)}_{1s}(j+1) = c^{(j+1)-(s-2)}, \quad 2 \leq s \leq (j+1)+2 \quad \text{and:} \\ a^{(l)}_{1s}(j+1) = a^{(l)}_{1(s-(j+1))}, \quad (j+1)+2 < s \leq J+2 \text{ (if } j+1 < J). \end{cases}$$

At last, (V_j) was constructed in such a manner that each customer in queue 2 advances of one stage at each step of this sequence, so that:

$$Q_{V_{j+1}}(1, j+3) = Q_{V_j}(1, j+2) = q',$$

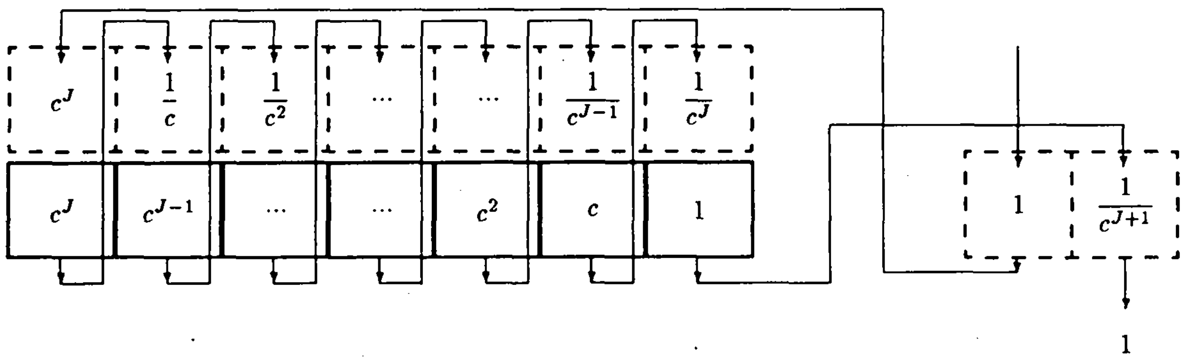

Figure 15 : Back to a reference state.

which completes the proof. □

It is easy to check that the network at time U is in the kind of state described in the above lemma. In consequence, if we set $V_0 = U$, then at time V_J queue 1 will be empty and queue 2 will be a homogeneous layer of composition $[c^J, c^{J-1}, ..., 1]$ (like in the initial state), with:

$$Q_{V_J}(1, J+2) = Q_U(1, 2).$$

Moreover, it is easy to check that V_J is in the form:

$$V_J = T(c)q.$$

Our last task consists in calculating $Q_U(1,2)$. In view of lemma 3.8 and the description of the state at time U below, we get:

$$\begin{aligned} Q_U(1,2) &= Q_U^{(1)}(1,2) + \sum_{i=2}^{J+1} Q_U^{(i)}(1,2) + Q_U^{(J+2)}(1,2) \\ &= c\frac{(1-c^J-c^{J+1})(1-c^{J^2}) - c^J(1-c^J)}{(1-c^{J^2})(1-c^{J+1})}q + \sum_{i=2}^{J} c^i q + \frac{c^{J+1}}{1-c^{J+1}}q + \frac{c^{J+2}(1-c^J)}{(1-c^{J^2})(1-c^{J+1})}q \\ &= c\frac{(1-c^J-c^{J+1})(1-c^{J^2}) - c^J(1-c^J)}{(1-c^{J^2})(1-c^{J+1})}q + c^2\frac{1-c^{J-1}}{1-c}q + \frac{c^{J+1}}{1-c^{J+1}}q + \frac{c^{J+2}(1-c^J)}{(1-c^{J^2})(1-c^{J+1})}q \end{aligned}$$

which is in the form: $Q_U(1,2) = \theta(c)q$. Now an easy calculus leads to the following condition.

Condition (H3): c must be chosen such that $\theta(c) > 1$, which is equivalent to:

$$\boxed{c^{2J+1}\frac{1-c^{(J-1)J}}{1-c^{J^2}} + c^2\frac{(1-c^{J-1})(1-c^{J+1})}{(1-c)^2} > 1}.$$

We define:

$$\begin{aligned} f : \quad [1,+\infty[\times[0,1] &\to \mathbb{R} \\ (J,c) &\mapsto 1 - c^J - c^{J+1} - c^J\frac{1-c^J}{1-c^{J^2}}, \end{aligned}$$

and:

$$\begin{aligned} g : \quad [1,+\infty[\times[0,1] &\to \mathbb{R} \\ (J,c) &\mapsto c^{2J+1}\frac{1-c^{(J-1)J}}{1-c^{J^2}} + c^2\frac{(1-c^{J-1})(1-c^{J+1})}{(1-c)^2} - 1. \end{aligned}$$

The conditions (H1), (H2) and (H3) are equivalent to:

$$f(J,c) > 0 \quad \text{and:} \quad g(J,c) > 0.$$

For any $J \geq 1$, $f(J,.)$ and $g(J,.)$ are continuous functions on $[0,1]$ and:

$$f(J,0) = 1,\ f(J,1) = -1 - \frac{1}{J}, \quad \text{and: } g(J,0) = -1,\ g(J,1) = J^2 - 1 - \frac{1}{J} > 0 \quad \text{for } J \geq 2.$$

Then for $J \geq 2$, we may define:

$$v_J = \inf\{v \in [0,1]/f(J,v) \leq 0\}, \quad u_J = \sup\{u \in [0,1]/g(J,u) \leq 0\},$$

and we have $f(J,c) > 0$ for $c \in [0, v_J[$, and $g(J,c) > 0$ for $c \in]u_J, 1]$. It is thus clear that Theorem 3.4 will be proven when we will have shown the following lemma.

Lemma 3.11
For $J \geq 3$, we have: $\frac{1}{2} < u_J < v_J < 1$, and $(u_J)_{J\in\mathbb{N}}$ (resp. $(v_J)_{J\in\mathbb{N}}$) is a non-increasing (resp. a non-decreasing) sequence converging to $1/2$ (resp. to 1).

Let us define:

$$\begin{array}{rll} \phi: & [1,+\infty[\times[0,1]\times[0,1] & \to \mathbb{R} \\ & (J,c,x) & \mapsto 1-(1+c)x - x\dfrac{1-x}{1-x^J}. \end{array}$$

We have: $f(J,c) = \phi(J,c,c^J)$, and then:

$$\begin{aligned} \frac{\partial f}{\partial J}(J,c) &= \frac{\partial \phi}{\partial J}(J,c,c^J) + \ln(c)c^J\frac{\partial \phi}{\partial x}(J,c,c^J) \\ &= -c^J(1-c^J)\frac{\ln(c^J)c^{J^2}}{(1-c^{J^2})^2} + \ln(c)c^J\left[-(1+c) - \frac{1-c^J}{1-c^{J^2}} - c^J\frac{-(1-c^{J^2}) + (1-c^J)Jc^{(J-1)J}}{(1-c^{J^2})^2}\right] \\ &= -c^J\ln c\left[1+c+2\frac{1-c^J}{1-c^{J^2}}\left(1+J\frac{c^{J^2}}{1-c^{J^2}}\right) - \frac{1}{1-c^{J^2}}\right] \\ &\geq -c^J\ln c\frac{c^{J^2}}{1-c^{J^2}}\left(-1+2J\frac{1-c^J}{1-c^{J^2}}\right) \quad \text{with:} \end{aligned}$$

$$\frac{1-c^{J^2}}{1-c^J} = \sum_{i=0}^{J-1} c^{iJ} \leq J, \quad \text{and then:}$$

$$\frac{\partial f}{\partial J}(J,c) \geq -c^J\ln c\frac{c^{J^2}}{1-c^{J^2}}(-1+2) = -c^J\ln c\frac{c^{J^2}}{1-c^{J^2}} \geq 0.$$

In consequence, (v_J) is a non-decreasing sequence; since it is upper bounded by 1, it has a limit $v \leq 1$. It is easy to check that when $J \to +\infty$, $f(J,.)$ converges pointwise (on $[0,1[$) to a limit function f which is defined by:

$$f(c) = 1 \text{ for } c \in [0,1[.$$

If $v < 1$, f is continuous on $[0,v]$, and Dini's theorem (see for example [5]) ensures us that the sequence $f(J,.)$ converges uniformly toward f on this compact interval. Hence we obviously have:

$$f(J,v_J) = 0 \to f(v) = 1, \quad \text{which is absurd.}$$

In consequence, $v = 1$.

Similarly we define:

$$\begin{aligned}\gamma: \quad [1,+\infty[\times[0,1]\times[0,1] &\to \mathbb{R}\\ (J,c,x) &\mapsto \frac{x^2}{c}\frac{1-c^2x^{J-2}}{1-cx^{J-1}}+\frac{(c^2-x)(1-x)}{(1-c)^2}-1.\end{aligned}$$

We have: $g(J,c)=\gamma(J,c,c^{J+1})$, and then trite calculations show that:

$$\begin{aligned}\frac{\partial g}{\partial J}(J,c) &= \frac{\partial \gamma}{\partial J}(J,c,c^{J+1})+\ln(c)c^{J+1}\frac{\partial \gamma}{\partial x}(J,c,c^{J+1})\\ &= -c^{J+1}\ln c\left[\frac{2Jc^{J^2}(1-c^J)-c^{J^2}(1-c^{J^2})}{(1-c^{J^2})^2}+\frac{2(c^{J^2}-c^J)}{1-c^{J^2}}+\frac{(1-c)^2+2c(1-c^J)}{(1-c)^2}\right]\\ &\geq -c^{J+1}\ln c\left[\frac{2c(1-c^J)}{(1-c)^2}-\frac{2c^J-1}{1-c^{J^2}}\right] \quad \text{with:}\end{aligned}$$

$$2c\left(\frac{1-c^J}{1-c}\right)^2\frac{1-c^{J^2}}{1-c^J}\geq 2c\geq 2c^J-1, \quad \text{and then:}$$

$$\frac{\partial g}{\partial J}(J,c)\geq 0.$$

In consequence, (u_J) is a non-increasing sequence; since it is lower bounded by 0, it has a limit $u\geq 0$.

It is easy to check that when $J\to+\infty$, $g(J,.)$ converges pointwise (on $[0,1[$) to a limit function g which is defined by:

$$g(c)=\frac{c^2}{(1-c)^2}-1, \text{ for } c\in[0,1[.$$

Since g is continuous on the compact interval $[0,u_2]$, we may apply Dini's theorem again to get that:

$$g(u)=\lim_{J\to+\infty}g(J,u_J)=0,$$

and then $u=1/2$.

The proof that $u_3<v_3$ is left to the reader ($f(3,.)$ and $g(3,.)$ are monotonous functions). □

Remark 3.12

- *Instead of invoking Dini's theorem, we could have used that: $u\mapsto u^n$ converges uniformly to 0 on $[0,x]$ for $x<1$, to make a direct proof of the uniform convergence of $f(J,.)$ and $g(J,.)$ on $[0,x]$.*
- *For numerical applications, one can check that:*

  $$u_3<0.557 \quad \textit{and} \quad v_3>0.749,$$

  and then the interval $[0.557, 0.749]$ is included in the domain of instability (for c) for all $J\geq 3$.

4 Conclusion.

In this paper, we considered fluid Bramson networks in the configuration (15). We identified a special class of states called reference states. We proved that for $J\geq 3$, there are values of c such that when the fluid network starts from a reference state, it eventually comes back to a larger reference state, and so on. This analysis was made in terms of homogeneous layers, which correspond to a piecewise linear evolution of the system.

For $J = 2$, this analysis fails but it is possible to find similar, unstable cycles (with the same reference state) in the following configuration:

$$\rho_{12} = c, \quad \rho_{1(J+3)} = c' \geq c, \quad \rho_{11} = \rho_{1s} = 0,. \quad 3 \leq s \leq J+2,$$

for some special values of c and c'. For $J = 1$, we conjecture that the network is stable under the usual conditions, but we have not been able to prove this result for the moment.

We think that unstable cycles still exist even in the case of small but non-null services, but in this case we may lose the description in terms of homogeneous layers. Notice anyway that our unstable cycle in configuration (15) can be approached as the limit of a fluid model with non-null service times. The natural way to deduce the transience of the stochastic model from the existence of unstable cycles would be to prove a kind of large deviation principle around our fluid trajectories (and here the fact that these trajectories are the unique solutions of the fluid equations governing the behaviour of the fluid model will of course be crucial). Some contractive property similar to that obtained by Seidman in [10] may be necessary in order to control the deviation of the stochastic model from the fluid one. We conjecture that the intervals of fluid instability that we identified in the configuration (15) are included in the domain of transience of the stochastic model for sufficiently small δ; the simulations that we did seem to agree with this conjecture.

Even if we limit the scope of our study to the fluid model, finding the exact conditions under which the network empties or not seems to be out of reach for the moment. We can only claim that for $J = +\infty$ and in the configuration (15), we have the exact domain of stability. As for the question of finding the exact stability conditions of the stochastic Bramson networks, it is far beyond our current capacity.

Finally, regardless of the conditions of stability, several questions about fluid models of multiclass queueing networks remain open: except, say, on the frontier of the domain of stability, is it true that the fluid limit model of an ergodic network always empties ? Does a transient network under the usual conditions always exhibit unstable cycles in its associated fluid model ? This should be the matter of future work.

References

[1] Botvitch, D., and Zamyatin, A. Ergodicity of conservative communication networks. Rapport de recherche 1772, INRIA, Domaine de Voluceau, Rocquencourt B.P. 105 78153 Le Chesnay Cedex France, October 1992.

[2] Bramson, M. Instability of FIFO queueing networks. Preprint, 1993.

[3] Dai, J. On the positive Harris recurrence for multiclass queueing networks: a unified approach via fluid limit models. Preprint, 1993.

[4] Dai, J., and Weiss, G. Stability and instability of fluid models for certain re-entrant lines. Preprint, February 1994.

[5] Dieudonné, J. *Éléments d'analyse: 1) Fondements de l'analyse moderne.* Gauthier-Villars (Paris), 1968.

[6] Dumas, V. Harris ergodicity of a multiclass queueing network via its associated fluid limit model. Rapport de recherche, INRIA, December 1993.

[7] KUMAR, P., AND SEIDMAN, T. I. Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. *IEEE Transactions on Automatic Control 35*, 3 (March 1990), 289–298.

[8] LU, S., AND KUMAR, P. Distributed scheduling based on due dates and buffer priorities. *IEEE Transactions on Automatic Control 36*, 12 (December 1991), 1406–1416.

[9] RYBKO, A., AND STOLYAR, A. Ergodicity of stochastic processes describing the operation of open queueing networks. *Problemy Peredachi Informatsii 28*, 3 (July-September 1992), 3–26.

[10] SEIDMAN, T. 'First Come, First Served' can be unstable ! Preprint, September 1993.

[11] V.A.MALYSHEV, AND M.V.MENSHIKOV. Ergodicity, continuity, and analyticity of countable Markov chains. *Trans. Moskow Math. Soc. 1* (1981), 1–47.

Les rapports de recherche de l'INRIA sont disponibles en format postscript sous ftp.inria.fr (192.93.2.54)

si vous n'avez pas d'accès ftp la forme papier peut être commandée par mail :
e-mail : dif.gesdif@inria.fr
(n'oubliez pas de mentionner votre adresse postale).

par courrier :
Centre de Diffusion
INRIA
BP 105 - 78153 Le Chesnay Cedex (FRANCE)

INRIA research reports are available in postscript format ftp.inria.fr (192.93.2.54)

if you haven't access by ftp we recommend ordering them by e-mail :
e-mail : dif.gesdif@inria.fr
(don't forget to mention your postal address).

by mail :
Centre de Diffusion
INRIA
BP 105 - 78153 Le Chesnay Cedex (FRANCE)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine - Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes - IRISA, Campus universitaire de Beaulieu 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)
Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399

RR_2318